THE ROLE OF SPT4/5 AND THE SEARCH FOR ANTITERMINATION COMPLEXES IN ARCHAEA

Daniel Joseph Fielden

A thesis submitted for the degree of Doctor of Philosophy

November 2014

0.1 Declaration of Authorship

I, Daniel Joseph Fielden confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

0.2 Abstract

Spt4/5 and its bacterial homologue NusG are the only known universally conserved RNAP-associated transcription elongation factors. In the hyperthermophilic archaeon *Methanocaldococcus jannaschii*, Spt5 comprises an N-terminal NGN domain and a C-terminal KOW domain, and is bound at its NGN domain by Spt4. NusG and Spt5 increase the processivity of RNAP by binding to the RNAP clamp via the NGN domain. This maintains the RNAP clamp in a closed conformation, thereby enabling RNAP to remain bound to the template DNA. The NusG KOW domain interacts with ribosomes, thereby coupling transcription to translation.

The functions of Spt4/5 in archaea are less well characterised. The work contained within this thesis demonstrates that in the context of *M. jannaschii* cell extract, Spt4/5 is found in the same fractions as ribosomes and RNAP, and therefore has the potential to couple transcription and translation. Furthermore, data obtained by microscale thermophoresis suggests that the KOW domain of Spt5 interacts with purified ribosomes.

Electron paramagnetic resonance was performed on Spt4/5, demonstrating that Spt5 is conformationally flexible, and that the presence of Spt4 restricts its mobility. Limited proteolysis and thermofluor assays support the notion that Spt4 stabilises the Spt5 NGN domain.

In *E. coli*, NusA binds to RNAP as a component of the antitermination complex, along with NusG, NusB, and NusE. This enables RNAP to enter a pause and termination-resistant state. *M. jannaschii* NusA consists of two KH domains. Mutational analysis identified the contribution of the two KH domains to RNA binding and identified additional residues involved in the interaction. Archaeal NusA does not coelute with RNAP, raising the possibility that archaeal NusA does not have antitermination functions.

In summary this thesis argues that Spt4/5 likely couples transcription and translation in archaea and indicates that archaeal NusA binds to RNA via a novel binding site.

0.3 Acknowledgements

I would like to express my sincere thanks to:

- My supervisor Professor Finn Werner for allowing me to undertake this exciting project within his lab. I am very grateful for his unfailing support during these last four years, for his guidance and for teaching me so much.

- All the members of the RNAP lab at UCL past and present. It has been a pleasure to work with you all. I would like to thank Fabian Blombach for his helpful advice on experiments and for finding us nice places to go for lunch. I would especially like to thank Tina Daviter for all the help and guidance with the MST and thermofluor experiments and for preparing ribosomes. Thank you to Thomas Fouqueau for making sure we get to the pub and have fun out of the lab. I am very grateful to Carol Sheppard for the critical reading of this thesis during its preparation. I would also like to thank Kathy Smollett for experimental advice and for all of the delicious baking.

- Dr Chris Kay for being my co-supervisor and for guidance throughout the project, and Dr Enrico Salvadori for trouble shooting about EPR, and for help with recording spectra.

- Dr Alethea Tabor for being my thesis committee chair.

- Professor Andres Ramos for guidance and advice regarding the NMR experiments.

- Dr Mark Pfuhl for allowing me to use his MST machine.

- Dr James Dicken for proofreading this thesis.

- I would especially like to thank my mum and grandma for their continued love and support and for believing in me.

Contents

List of Figures

List of Tables

0.4 Abbreviations

2-P ESEEM	Two-Pulse Electron Spin Echo Envelope Modulation
A	Adenine
ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
APS	Ammonium Persulfate
bp	Base Pair
BRE	B-recognition Element
BSA	Bovine Serum Albumin
C	Cytosine
cDNA	Complementary DNA
ChIP	Chromatin Immunopreciptation
CLIP	Crosslinking and Immunopreciptation
CryoEM	Cryo-electron Microscopy
CTR	C-terminal Repeats
CW	Continuous Wave
Da	Daltons
DEER	Double Electron Electron Resonance
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleoside Triphosphate
ds	Double-stranded
DSIF	DRB-sensitivity Inducing Factor
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
EMSA	Electrophoretic Mobility Shift Assay
EPR	Electron Paramagnetic Resonance
FPLC	Fast Performance Liquid Chromatography
G	Gauss
G	Guanine
g	Earth's Gravitational Acceleration
g	Grams
GHz	Gigahertz
GTP	Guanosine Triphosphate
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl-β-D-1-thiogalactopyranoside

K	Kelvin
kb	Kilobases
kDa	Kilodaltons
KOW	Kyrpides-Ouzounis-Woese
L	Litres
LB	Lysogeny Broth
M	Molar
MHz	Megahertz
mg	Milligrams
ml	Millilitres
mM	Millimolar
mRNA	Messenger RNA
MST	Microscale Thermophoresis
MW	Molecular Weight
nm	Nanometres
nM	Nanomolar
NELF	Negative Elongation Factor
NGN	NusG N-terminal Domain
NMR	Nuclear Magnetic Resonance
NPS	Nano Positioning System
NTP	Nucleoside Triphosphate
nt	Nucleotide
Nut	N-Utilisation Site
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
P. furiosus	*Pyrococcus furiosus*
Pi	Inorganic Phosphate
PPi	Pyrophosphate
rcf	Relative Centrifugal Force
RFU	Relative Fluorescent Units
RNA	Ribonucleic Acid
RNAP	RNA Polymerase
rRNA	Ribosomal RNA
RT PCR	Real Time Polymerase Chain Reaction
S	Svedberg Unit
S. cerevisiae	*Saccharomyces cerevisiae*

SDS	Sodium Dodecyl Sulfate
SDSL	Site-directed Spin Labelling
SIA	Scaffold-independent Analysis
ss	Single-stranded
T	Thymine
TAE	Tris Acetate EDTA
TAE	Tris Borate EDTA
Taq	*Thermus aquaticus*
TBP	TATA-binding Protein
TCEP	Tris-(2-carboxyethyl)phosphine
TEMED	Tetramethylethylenediamine
TFB	Transcription Factor B
TFE	Transcription Factor E
TFS	Transcription Factor S
Tm	Melting Temperature
TNB	3-thio-6-nitrobenzoate
tRNA	Transfer RNA
UV	Ultra Violet
μ**g**	Micrograms
μ**l**	Microlitres
μ**M**	Micromolar
WT	Wild Type

Chapter 1

Introduction

1.1 Archaea

Archaea were first identified as a class of organism distinct from both bacteria and eukaryotes in 1977, based on the sequencing of 16S ribosomal RNA (Woese and Fox, 1977), leading to their classification as a separate domain of life (Woese et al., 1990). Archaea are single-celled organisms that can range from 0.1 to over 15 μm in length. Archaea exhibit a diverse range of morphologies and some archaeal species form aggregates that can be 200 μm in length. Archaea are often associated with living in extreme conditions (Chaban et al., 2006). They have been observed to grow at temperatures as high as 121°C (Kashefi and Lovley, 2003), and have also been isolated from cold environments such as Antarctica (Cavicchioli, 2006). The archaeal genus *Picrophilus* has been observed to grow at a pH very close to 0 (Schleper et al., 1995). However archaea also inhabit fresh water and have been isolated from humans within the large intestine (Miller and Wolin, 1982). One distinguishing feature of archaea is their cell membranes, which comprise mainly glycerol-ether lipids (Gambacorta et al., 1995). This is in contrast to bacteria and eukaryotes whose membranes are glycerol-ester based. In terms of metabolic pathways, archaea are more closely related to bacteria than to eukaryotes. The organisation of their genomes is also closer to bacteria than to eukaryotes, being small and circular with genes often arranged in operons. However, archaea are more reminiscent to eukaryotes with respect to information processing. Archaeal DNA replication, transcription, and translation all closely resemble their eukaryotic counterparts.

The precise origins of archaea is still under dispute, with two main arguments proposed (Gribaldo et al., 2010). The "three domains" hypothesis argues that bacteria, archaea, and eukaryotes are three distinct lineages, each of which has its own common ancestor. Within the three domain hypothesis, archaea and eukaryotes are "sister lineages" which share an ancestor prior to their split (Figure 1.1A). In contrast, the "two domains", or eocyte, hypothesis argues that there are only two distinct lineages: bacteria and archaea (Figure 1.1B). According to the eocyte hypothesis, eukaryotes arose as the result of a symbiotic relationship between bacteria and crenarchaea. Thus within the two domains hypothesis eukaryotes are not a distinct lineage. (Williams et al., 2013).

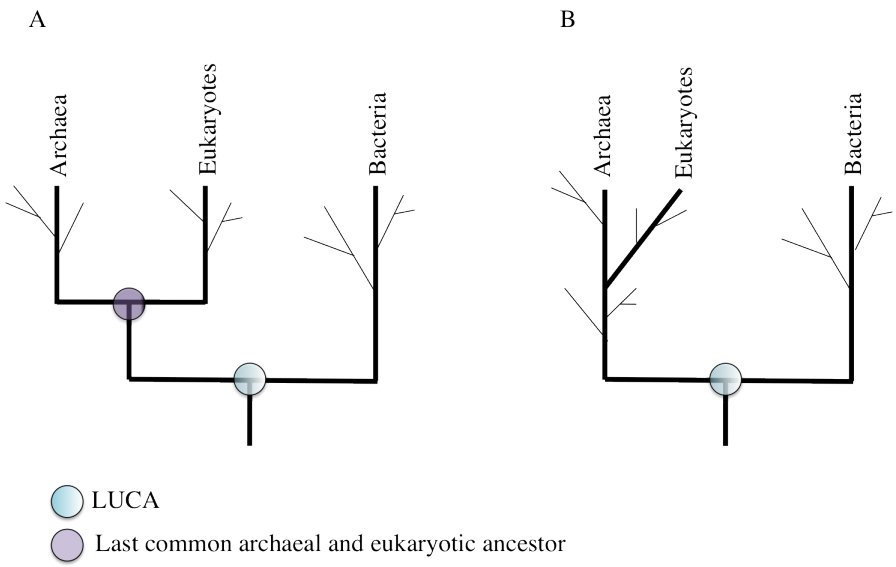


Figure 1.1: Two hypotheses of the tree of life. A) In the traditional three-domain model there are three distinct branches: bacteria, archaea and eukaryotes. Archaea and eukaryotes split after the separation from bacteria. All archaea and eukaryotes can be traced back to a common archaeal/eukaryotic ancestor. B) In the eocyte hypothesis there are only two distinct branches: bacteria and archaea. Eukaryotes arose through a symbiotic relationship between crenarchaea and bacteria. Thus archaea and eukaryotes do not all share a common archaeal/eukaryotic ancestor.

The archaeal domain can be subdivided into distinct phyla. The two main archaeal phyla are the crenarchaea and the euryarchaea. The crenarchaea are more closely related to eukaryotes than the euryarchaea, and include the well-studied species *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius*. The euryarchaea are less closely related to the eukaryotes and include the methanogens and haloarchaea. Among the well-studied euryarchaea are *Methanocaldococcus jannaschii* and *Pyrococcus furiosus*. Recently, several additional phyla of archaea have been proposed. These include the korarchaea which have properties in common with both the crenarchaea and the euryarchaea (Elkins et al., 2008), the thaumarchaea which are closely related to the crenarchaea but which are typically mesophilic (Brochier-Armanet et al., 2008), and the nanoarchaea, of which there is currently only one characterised member, *Nanoarchaeum equitans*. *N. equitans* has a genome of only 480 kb, making it the smallest cellular genome to be sequenced (Brochier et al., 2005).

The work in this thesis seeks to enhance knowledge of transcription in the archaeon *M. jannaschii*.

M. jannaschii is a euryarchaea that grows in hydrothermal vents at approximately 80°C, under high pressure and anaerobic conditions (Figure 1.2)[1]. *M. jannaschii* is a methanogen, which generates energy through the reduction of CO_2 to methane (Zhu et al., 2004). It was the first archaeal species to have its complete genome sequenced (Bult et al., 1996). This identified that is has one large chromosome of 1.66 mega basepairs, a large circular extra chromosome of 58 kilo basepairs, and a small circular extra chromosome of 16.5 kilo basepairs (Bult et al., 1996). Interestingly the *M. jannaschii* genome is AT rich with only 30% GC content. However the proteins have evolved to life at high temperatures. Compared to homologues from mesophilic members of the *Methanococcus* phylum, *M. jannaschii* proteins contain more residues, with a greater proportion being hydrophobic or charged (Haney et al., 1999).

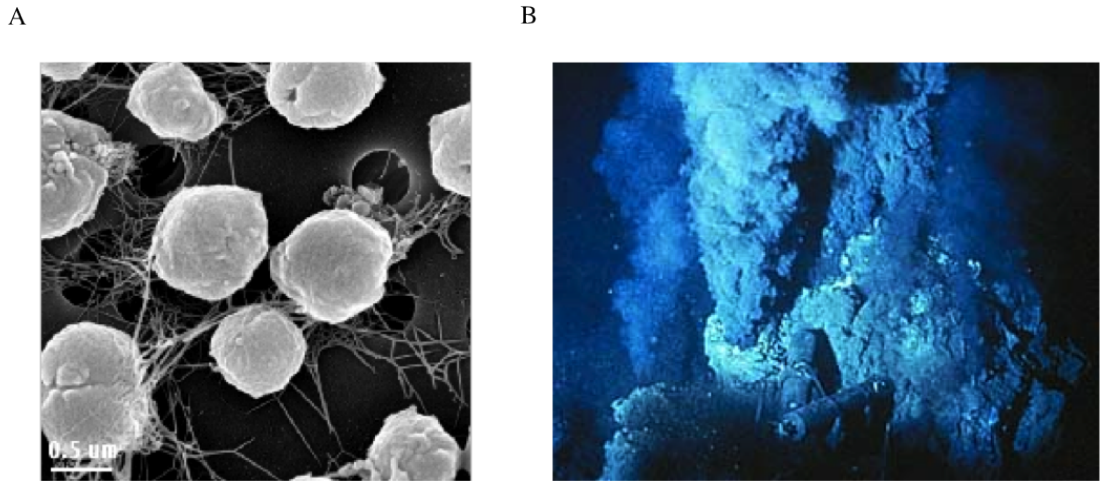


Figure 1.2: A) *M. jannaschii* cells viewed under an electron microscope. *M. jannaschii* has an uneven spherical shape and has flagella to enable motility. B) *M. jannaschii* grows at high temperature and pressure in hydrothermal vents.

1.2 Transcription Overview

Transcription is the synthesis of an RNA polymer from monomeric nucleoside triphosphates (NTPs), using a DNA template. The enzyme that carries out this process is called RNA polymerase (RNAP). This essential enzyme is found in all forms of life (Werner and Grohmann, 2011). Transcription can be divided into three key phases: initiation, elongation and termina-

[1]Figure 1.2 photos from https://microbewiki.kenyon.edu/index.php/Methanococcus_jannaschii

tion. During initiation the RNAP binds to the promoter and the DNA is melted, exposing the template DNA bases. During elongation the RNAP moves along the DNA template catalysing the addition of nucleotides to the 3′ end of the nascent RNA chain. Finally the RNAP must dissociate from the template DNA and release the transcript during termination. After termination RNAP is free to initiate another round of transcription. Thus transcription is a cyclic process.

1.3 RNA Polymerase Structure and Function

RNAPs from all three domains of life show a high degree of structural similarity with a characteristic crab claw appearance (Campbell et al., 2001; Cramer et al., 2001; Hirata et al., 2008; Werner and Grohmann, 2011) (Figure 1.3A). The catalytic subunits across the three domains all contain a double-psi-β-barrel motif, suggesting that they originate from a common ancestor (Werner and Grohmann, 2011). Bacterial RNAP is the simplest, containing five subunits: α_1, α_2, β, β', and ω (Figure 1.3B). The β and β' subunits make up the catalytic core and comprise approximately 70% of the RNAP mass. The two α subunits are required for assembly of the multi-subunit complex (Werner and Weinzierl, 2002). The ω subunit has a stabilising effect on RNAP (Mukherjee and Chatterji, 1997). This core structure is conserved in archaeal and eukaryotic RNAPs, which have homologues of all bacterial RNAP subunits plus a number of additional subunits.

In archaea the two catalytic subunits (Rpo1 and Rpo2) are closely related to the bacterial β' and β subunits. In many archaea the genes encoding the catalytic subunits have undergone a gene splitting process. In *P. furiosus* the β' homologue is encoded by Rpo1 and Rpo1′. In *M. jannaschii* the β homologue has also split into two polypeptides, Rpo2 and Rpo2′ (Werner, 2007). The catalytic subunits contain the trigger loop and bridge helix (for substrate selection and RNAP translocation respectively), two Mg^{2+} ions (one permanently bound, the other enters with NTPs and leaves with PP_i) and the clamp which encloses the DNA during transcription (Werner and Grohmann, 2011). Archaeal RNAP contains four subunits that make up the assembly platform (Rpo3, Rpo11, Rpo10 and Rpo12) in an analogous manner to the two α subunits of bacterial RNAP. They form a stable heterotetrameric complex *in vitro* (Werner et al., 2000) and all four are required for assembly of the polymerase (Werner and Weinzierl, 2002). Archaeal

RNAP also contains a number of auxiliary subunits with additional functions (Rpo4, Rpo5, Rpo6, and Rpo7) (Figure 1.3C). Rpo4 and Rpo7 form a stable sub-complex together and are non-reversibly incorporated into the RNAP (Grohmann et al., 2009). Subunit Rpo6 is required for the incorporation of Rpo4/7. When it is deleted from RNAP, the RNAP is unable to form a stable complex with free Rpo4/7. Rpo5 contacts Rpo1 forming the lower jaw and is important for open complex formation (Grünberg et al., 2010).

Not all archaeal RNAPs have an identical subunit arrangement. Hyperthermophilic crenarchaea and korarchaea also contain a subunit called Rpo8 which is not present in euryarchaea (Grünberg et al., 2010; Koonin et al., 2007). Rpo8 is homologous to the eukaryotic RNAPII subunit RPB8. Crenarchaeal RNAP also contains a subunit called Rpo13 which has no known euryarchaeal nor eukaryotic homologue. Structural data suggests that Rpo13 stabilises the downstream DNA-RNAP interaction (Wojtas et al., 2012).

The most striking difference between bacterial and archaeal/eukaryotic RNAP is the presence of the stalk domain that is composed of subunits Rpo4/7 in archaea and RPB4/7 in eukaryotes. In *M. jannaschii*, the Rpo4/7 subcomplex has multiple functions. During transcription initiation it promotes formation of the open complex and is necessary for the initiation factor TFE to stimulate transcription (Ouhammouch et al., 2004). During elongation it enhances the processivity of RNAP (Hirtreiter et al., 2010b) and it also stimulates termination (Hirtreiter et al., 2010b). The Rpo4/7 stalk guides the emerging transcript away from the RNAP (Hirtreiter et al., 2010b). In eukaryotes, when RPB4/7 binds to RNAP, the RNA-interacting tip undergoes a loop-to-helix conformational change (Armache et al., 2005). However, solution studies in *M. jannaschii* have indicated that Rpo4/7 binding does not result in any large scale conformational change within Rpo4/7 (Grohmann et al., 2010).

Bacteria and archaea have one RNAP which is responsible for the expression of all RNAs, whereas eukaryotes have three to five distinct RNAPs (I, II, III, IV, and V) which share a conserved core. RNAPII synthesises mRNAs that will be transcribed into proteins (Woychik and Hampsey, 2002). Of the eukaryotic RNAPs, RNAPII is most similar to that found in archaea (Figure 1.3D). The other eukaryotic RNAPs are more complex and have gained additional sub-

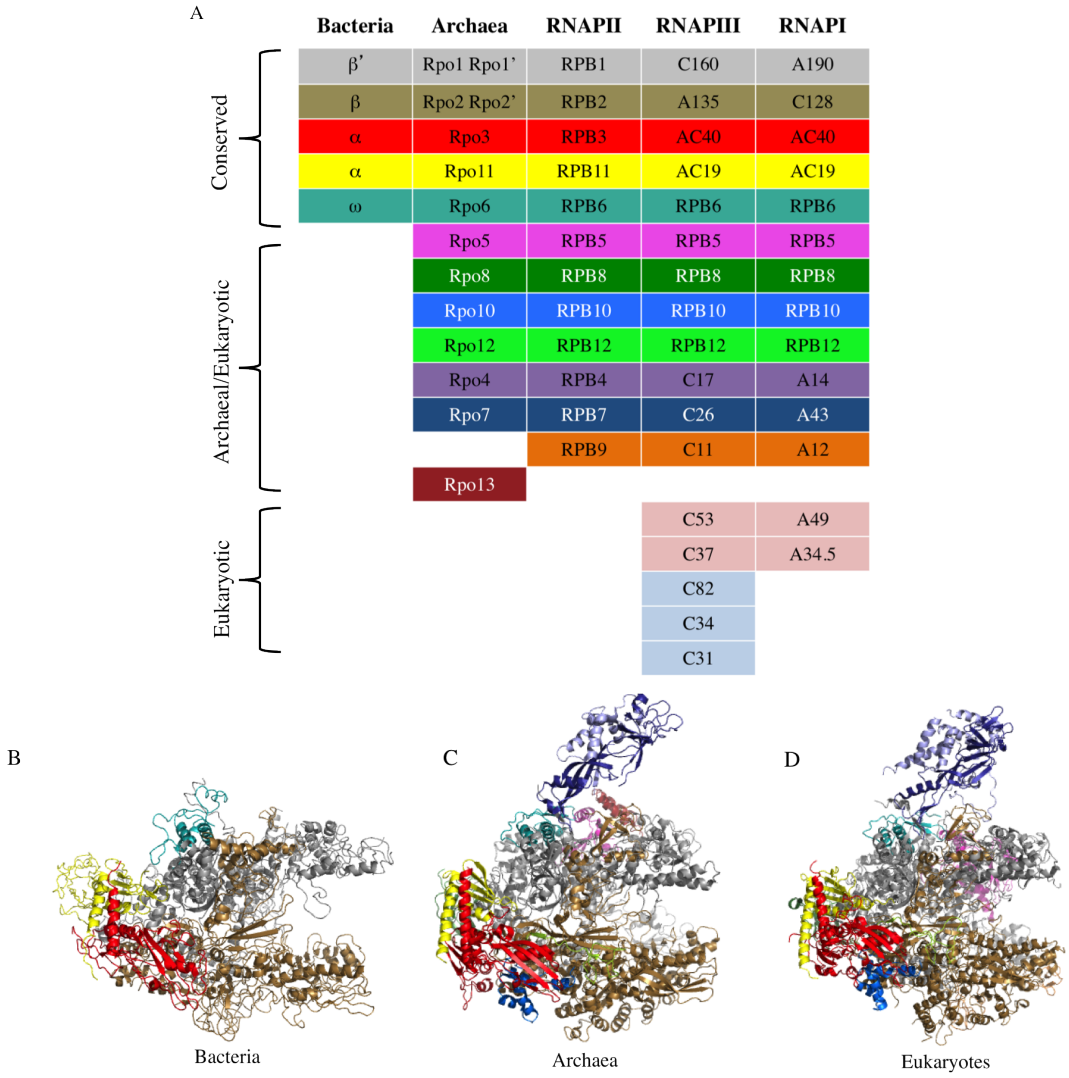


Figure 1.3: Subunit and structural conservation of RNAPs across the three domains of life. A) Bacteria, archaea and eukaryotes have a common core subunit composition. Archaeal and eukaryotic RNAPs are more complex, with a number of additional subunits. B, C, and D) All RNAPs have a common core 'crabclaw' structure with the two catalytic subunits forming the 'pincers' (grey and sand). The most obvious difference between bacterial and archaeal/eukaryotic RNAP is the addition of the Rpo4/7 stalk. Bacterial RNAP: *Thermus aquaticus* (PDB 1HQM), archaeal RNAP: *Sulfolobus shibate* (PDB 2Y0S), eukaryotic RNAPII: *Saccharomyces cerevisiae* (PDB 1Y1V).

units to perform specific functions within the cell. RNAPI synthesises the precursors to the 18S, 28S, and 5.8S ribosomal RNAs (rRNA) (Grummt, 1998) within the nucleolus (Grummt, 2003), comprising over 50% of cellular transcriptional output (Russell and Zomerdijk, 2006). RNAPIII is mainly involved in the synthesis of tRNAs and the 5S rRNA (Schramm and Hernandez, 2002;

Willis, 2005). RNAPIV and RNAPV are found only in plants (Herr et al., 2005) and are involved in gene silencing through the synthesis of siRNAS and DNA methylation (Kanno et al., 2005; Onodera et al., 2005).

1.4 Initiation and the Role of Basal Transcription Factors

1.4.1 Initiation in Bacteria

The first stage of transcription is called initiation, during which RNAP binds to the promoter and opens the double-stranded DNA, forming the transcription bubble. Unlike DNA polymerase, RNAP does not require a DNA or RNA primer to initiate transcription. RNAPs from all three domains of life require additional factors to initiate transcription in a promoter-dependent and start-site-specific fashion. However, these vary between bacterial and archaeal/eukaryotic RNAPs. Bacterial RNAP requires only a sigma factor to initiate transcription in a sequence-specific manner. σ factors recognise specific gene promoters (Paget et al., 2003) through association with the -10 and -35 regions (Campbell et al., 2008) and promote open complex formation (Browning and Busby, 2004). In bacteria, different σ factors enable gene-specific regulation of transcription. In *E. coli*, the main class of σ factor, σ^{70}, regulates expression of "housekeeping" genes, which are required for normal growth. σ^{H} is σ^{70}-like and promotes expression of genes in response to heat shock (Browning and Busby, 2004). σ^{54} regulates the expression of genes in response to nitrogen starvation. σ^{54} is the major class of alternative σ factor and recognises regions -12 and -24. σ^{54} requires ATP hydrolysis from a hexameric AAA+ activator to melt the DNA, in contrast to σ^{70} (Buck et al., 2000).

1.4.2 Initiation in Eukaryotes

Eukaryotic RNAPII initiation requires a host of basal initiation factors, namely TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. The basal initiation factors regulate the expression of all genes, unlike the σ factors which enable gene-specific regulation. Initiation begins with the binding of TATA-binding protein (TBP), a component of the TFIID complex that contains an additional 13 proteins, to the TATA box. The TATA box is a region of DNA that is about 25 nu-

cleotides upstream from the transcription start site. TBP is a saddle-shaped protein that binds in the minor groove of DNA (Kim et al., 1993a,b). The binding of TBP to the TATA box at the promoter causes the DNA to bend 90° (Kim et al., 1993a,b). This enables TFIIB to bind. TFIIB is composed of two domains; an N-terminal Zn-ribbon domain and a C-terminal core domain which are connected by a flexible linker (Figure 1.4). The N-terminal domain recruits RNAP to the initiation complex through interaction with the RNAP dock domain, and the C-terminal domain binds to and stabilises the TBP-TATA complex (Bushnell et al., 2004). In addition, the core domain of TFIIB makes sequence-specific contacts to a region of DNA called the B-recognition element (BRE) (Lagrange et al., 1998). TFIIB stimulates abortive and full-length transcription through insertion of a flexible region, called the B finger, deep into the catalytic site of RNAP (Werner and Weinzierl, 2005). Furthermore, the TFIIB B reader facilitates the separation of the DNA-RNA duplex, guiding the RNA towards the exit channel (Sainsbury et al., 2013). After TFIIB has bound, a factor called TFIIA binds to the pre-initiation complex and stabilises it by interacting with TBP of the TFIID complex (Buratowski et al., 1989). TFIIA also acts as an anti-repressor, competing with repressors such as NC2 for binding to the pre-initiation complex (Xie et al., 2000). Next, RNAP which is in complex with TFIIF binds to the pre-initiation complex. TFIIF stabilises the pre-initiation complex. RNAP then recruits TFIIE, which facilitates template DNA melting (Holstege et al., 1995). Finally a complex called TFIIH binds. Separation of the duplex DNA to form the open complex requires energy from ATP hydrolysis. This is achieved by a subunit of TFIIH called XPB (Grünberg et al., 2012).

1.4.3 Initiation in Archaea

In archaea, initiation of transcription utilises just three basal factors: TBP, TFB, and TFE (Werner and Weinzierl, 2005). Initiation in archaea therefore represents a streamlined version of that found in eukaryotes. Initiation begins with the binding of TBP to the TATA box. In archaea TBP binds to the TATA box by itself, rather than as part of a large complex. This is followed by the binding of TFB. TBP and TFB are necessary and sufficient for initiation *in vitro*. The third factor, TFE consists of a winged-helix domain and a zinc ribbon domain, and is homologous to the N-terminus of eukaryotic TFIIEα (Bell et al., 2001). There is no complete structure available for TFE, although the archaeal winged-helix domain alone has been solved by

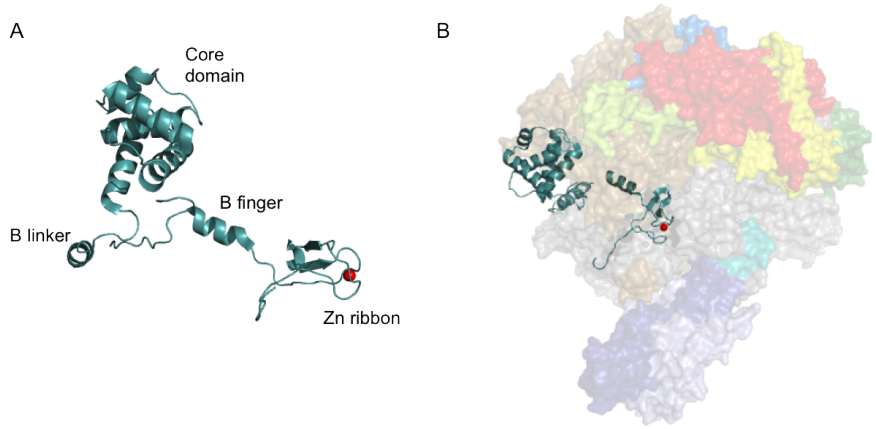


Figure 1.4: A) Crystal structure of TFIIB bound to RNAP (PDB 3K1F). TFIIB comprises an N-terminal Zn ribbon domain which binds RNAP, and a C-terminal core domain which binds to TBP. B) TFIIB binds to RNAP, inserting its B finger into the RNAP active centre.

X-ray crystallography (Meinhart et al., 2003) and the human zinc-ribbon domain from TFIIEα has been solved by NMR (Okuda et al., 2004), enabling the generation of a homology model (Grohmann et al., 2011). TFE binds to RNAP and stimulates the opening of the duplex DNA through its interaction with the upstream end of the partially-formed transcription bubble (Forget et al., 2004). In contrast to eukaryotes, no ATP hydrolysis is required for the formation of the transcription bubble in archaea.

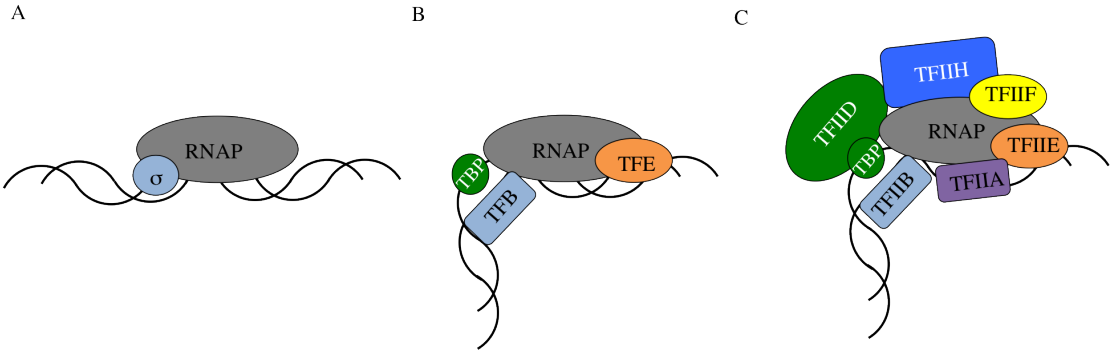


Figure 1.5: Comparison of the basal transcription machinery across the three domains of life. A) Bacterial RNAP requires only a σ factor to initiate transcription in a promoter-specific manner. Different σ factor classes provide gene-specificity. B) Archaea utilise three initiation factors called TBP, TFB and TFE. C) Eukaryotes require many basal factors to initiate transcription in a sequence-specific fashion. The same basal machinery is used for all genes.

Bacteria	Archaea	Eukaryotes (RNAPII)	Function
σ factors			Sequence-specific recruitment of RNAP to promoter, DNA strand separation, and promoter escape
	TBP	TBP (component of TFIID complex)	Recruits TF(II)B to promoter
	TFB	TFIIB	Recruits RNAP to promoter, transcription start site selection
	TFE	TFIIE	Stimulates DNA melting, stabilises open complex
		TFIIA	Stabilisation of pre-initiation complex and cancelling out repressors e.g. NC2
		TFIIF	Stabilises pre-initiation complex
		TFIIH	DNA melting

Table 1.1: Summary of basal transcription factors across the three domains of life and their function. Bacteria use just a sigma factor which binds the promoter, helps form the open complex, and facilitates promoter escape. Archaea have three basal initiation factors to perform these functions. Eukaryotes utilise many basal transcription factors.

After recruitment of the basal initiation factors, the open complex is formed in which about 14 basepairs of double-stranded DNA at the promoter are unwound. This exposes the nucleotide bases of the template strand, enabling their base-pairing with incoming nucleotides. RNAP then undergoes multiple rounds of abortive initiation during which small RNA products of 4-11 nucleotides in length are produced. In archaea and eukaryotes the short transcripts clash with the B-reader of TF(II)B. Similarly in bacteria, the transcript clashes with σ region 3.2 (Murakami et al., 2002). Data obtained by FRET indicates that during initiation RNAP draws the downstream template DNA into the active site, whilst the upstream DNA remains in a constant position with respect to RNAP (Kapanidis et al., 2006). This has been explained by a scrunching mechanism in which the unwinding of approximately one turn of DNA and the pulling of said DNA into the active site results in a stressed intermediate. The stress is relieved either by disengagement of RNAP with the downstream DNA (resulting in abortive initiation) or the escape of RNAP from the promoter (Revyakin et al., 2006). During promoter escape in archaea and eukaryotes the $3'$ end of the RNA transcript displaces the TF(II)B B-reader (Kostrewa et al., 2009).

1.5 General Mechanism of Transcription Elongation

1.5.1 The Nucleotide Addition Cycle

RNAP catalyses the formation of a phosphodiester bond between the 3′ terminus of a nascent RNA chain and an NTP using DNA as a template. The double-stranded DNA enters the RNAP via the DNA-entry channel, which is located between the two RNAP 'pincers' of the catalytic subunits. In order act as a template, double-stranded DNA must be progressively unwound to expose the DNA bases of the template strand and to enable their base-pairing with the NTP that is to be incorporated into the RNA chain. The incoming nucleotides enter the active site via a pore called the secondary channel which has dimensions of 15 x 20 Å. This is too small for the passage of double-stranded DNA (Vassylyev et al., 2007b). During elongation RNAP undergoes a repeated nucleotide addition cycle comprising of i) entry of the NTP into the RNAP active site which subsequently base-pairs with the template DNA at position i+1, ii) the RNAP-catalysed condensation reaction between the 3′ end of the nascent RNA chain which is in the i-site, and the NTP which is in the i+1 site, iii) the release of PP_i, and iv) translocation of RNAP one nucleotide along the DNA such that the newly-incorporated nucleotide is now in the i-site. Efficient gene expression depends on this cycle being rapidly repeated up to thousands of times.

1.5.2 Role of the Bridge Helix and Trigger Loop

The RNAP translocation mechanism relies on a mobile element of the β'/Rpo1 subunit called the bridge helix. The bridge helix has been observed in a straight conformation (Vassylyev et al., 2007a) and in a kinked conformation (Vassylyev et al., 2002) (Figure 1.6). Molecular dynamics studies suggest that there are two flexible hinges, called the N-terminal hinge and C-teminal hinge, within the bridge helix that enable it to adopt the kinked conformation (Weinzierl, 2010). The ability to undergo conformational changes is believed to be important in the nucleotide addition cycle. Mutations within the bridge helix that introduce a proline to destabilise the helix and to induce kinking also increase the activity of RNAP in transcription assays (Tan et al., 2008).

One hypothesised mechanism proposes that the bridge helix is in a straight conformation when

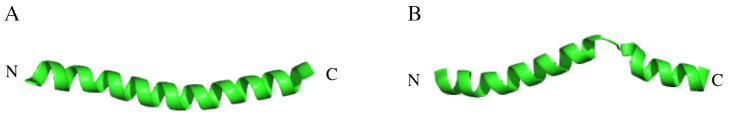


Figure 1.6: The bridge helix has been observed in A) straight and B) bent conformations. The interconversion between the straight and bent states is believed to be important for RNAP translocation during elongation. The bridge helix spans residues β' V1067 – H1103 in *Thermus thermophilus*, from which A) and B) are both taken (PDB 2O51 and 1IW7 respectively).

the incoming nucleotide enters the i+1 site and during catalysis. The bridge helix then undergoes a conformational change into the kinked state, translocating the nascent RNA such that the newly incorporated 3′ end moves from the i+1 site into the i site. The bridge helix then undergoes a conformational change back to the straight conformation, without further translocation of RNA, preparing RNAP for incorporation of the next NTP (Gnatt et al., 2001). RNAP translocates via a brownian-ratchet mechanism with only thermal fluctuations required for RNAP translocation (Bar-Nahum et al., 2005). RNAP oscillates forwards and backwards in a stochastic manner, fluctuating between the pre- and the post-translocated state. Upon NTP binding the equilibrium is shifted such that RNAP tends towards the post-translocated state. The RNAP oscillations have a distance of 3.4 Å, consistent with the distance between two nucleotides in B-form DNA (Abbondanzieri et al., 2005).

The Trigger Loop

Another important mobile element in the nucleotide addition cycle is the trigger loop, which is also a component of the catalytic β'/Rpo1 subunit. The trigger loop is required for selection of the correct nucleotide during elongation. In the absence of NTP it is disordered (Figure 1.7A). However, in the presence of the correct NTP, it undergoes a structural transition and makes contacts with the NTP ribose 3′ OH, the nucleotide base, and the β phosphate, resulting in the enclosure of the nucleotide (Wang et al., 2006) (Figure 1.7B). If a mismatched NTP enters the active site, the trigger loop will not enclose it, and catalysis will not occur. Mutations of the trigger loop decrease the fidelity of RNAP (Fouqueau et al., 2013; Kaplan et al., 2008). Furthermore, the trigger loop has a role in the hydrolysis of the 3′ nucleotide in backtracked RNAP (Yuzenkova and Zenkin, 2010).

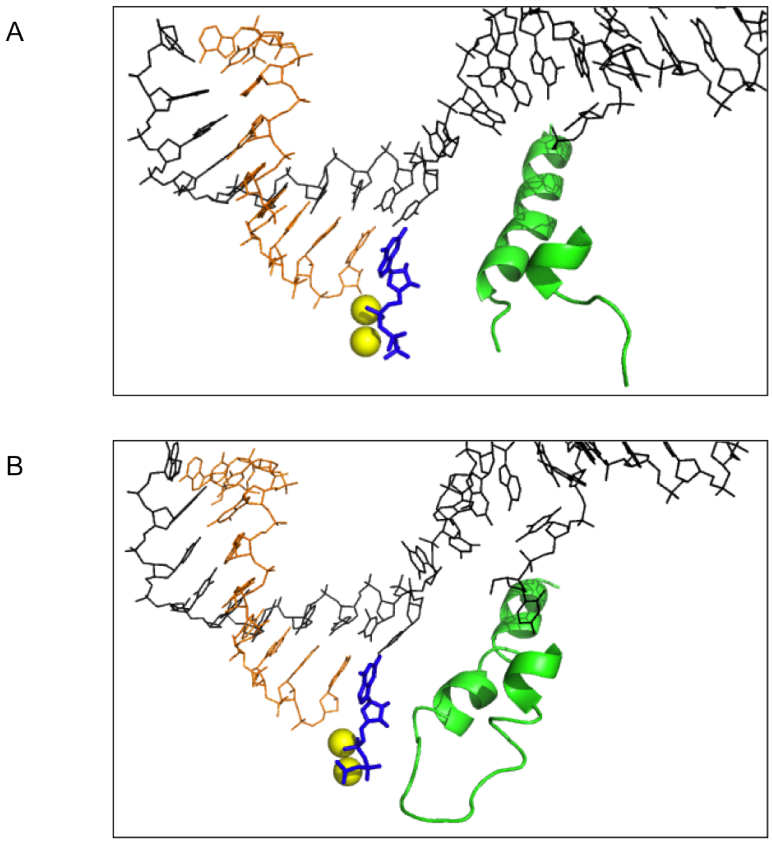


Figure 1.7: The trigger loop is involved in substrate selection. A) If no nucleotide or a non-cognate nucleotide is present in the active site, the trigger loop is disordered (PDB 2E2J). GMPCPP is shown in the active site. B) Upon binding of the correct nucleotide into the active site, the trigger loop undergoes a conformational change into its closed state where it encloses the nucleotide, enabling catalysis to occur (PDB 2E2H). GTP is shown in the active site. Green: trigger loop (*S. cerevisiae* RPB1 P1060 – L1101), black: DNA, orange: RNA, yellow spheres: Mg-1 and Mg-2, blue: nucleotide.

1.5.3 The Clamp

The clamp is a mobile element of the Rpo1 subunit that exists in an open conformation in free RNAP but which can rotate up to 30° about a switch region in yeast during transcription elongation (Gnatt et al., 2001). The closing of the clamp enables the stable association of the DNA and stabilises the elongation complex. In addition to the open and closed states, a collapsed clamp state has also been observed in free RNAP in which the clamp is even more closed than during elongation (Chakraborty et al., 2012). Within this state the DNA entry channel is too

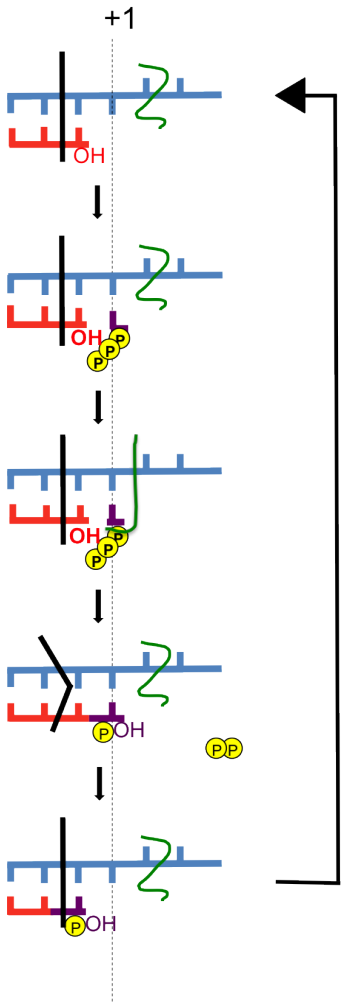


Figure 1.8: The nucleotide addition cycle begins after translocation with the 3′ terminus of the nascent RNA chain in the i position (red). An NTP (purple) enters the i+1 site and basepairs with the template DNA (blue). The trigger loop (green) then undergoes a conformational change, enclosing the nucleotide and catalysis occurs. The NTP is incorporated into the RNA chain and PP_i is released. The bridge helix (black) undergoes a conformational change, resulting in the translocation of the DNA-RNA hybrid within the active site so that the 3′ terminus is once again in the i site, enabling the cycle to begin again. Figure inspired by (Cramer et al., 2008).

narrow to enable the loading of DNA, and it is not therefore apparent whether the collapsed state has a biological function.

1.5.4 Additional Mobile Elements

During elongation the transcription bubble is maintained. The downstream DNA must continually be separated, whilst at the 5′ end of the transcription bubble the DNA-RNA hybrid must be melted and the template and non-template DNA strands must reanneal. A number of mobile elements within RNAP facilitate these functions. Fork loops 1 and 2 separate the template and non-template DNA strands at the downstream edge of the transcription bubble. The zipper functions in the reannealing of the two DNA strands. A region called the wall limits the length of the DNA-RNA hybrid to 8-9 basepairs and also causes the template strand to bend 90° in its exit path from RNAP. The rudder contributes to the stability of the elongation complex (Cramer et al., 2001; Gnatt et al., 2001; Kuznedelov et al., 2002).

1.5.5 Catalytic Mechanism of RNAP

Catalysis occurs by a conserved two-metal mechanism (Steitz, 1998) involving two Mg^{2+} ions that are coordinated by three evolutionary conserved aspartates (Zaychikov et al., 1996) of the absolutely conserved DXDGD motif. One of the Mg^{2+} ions, referred to as Mg-1, is permanently bound within the RNAP active site by the aspartates. The other Mg^{2+} ion, referred to as Mg-2, enters with every NTP, associated with the β and γ phosphates, and leaves with the liberated PP_i. Structural data suggests that the incoming NTP initially binds to a pre-insertion site, before being 'flipped' 180° into the insertion site, where it is appropriately aligned for incorporation into the transcript (Westover et al., 2004). The 3′ OH group at the 3′ terminus of the nascent RNA chain is coordinated by Mg-1, yielding an activated 3′ O^-. The activated 3′ terminus then nucleophilically attacks the α phosphate of the NTP in an SN2 reaction and an electron pair is subsequently transferred to the β phosphate resulting in the release of PP_i (Figure 1.9). RNAP also has exonuclease activity, for example when non-complementary nucleotides enter the active site (Sosunov et al., 2003).

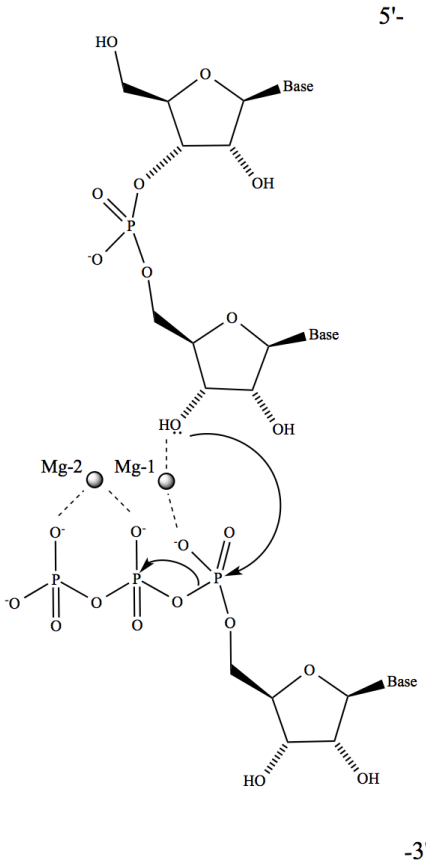


Figure 1.9: Catalytic mechanism of RNAP. The 3′ OH of the nucleotide in the i site is activated by Mg-1 and nucleophilically attacks the α phosphate of the incoming nucleotide. An electron pair is subsequently transferred to the β phosphate of the incoming nucleotide and PP_i is liberated.

1.5.6 The Role of RNA-Processing Factors

RNAP frequently pauses and backtracks along the DNA template (Werner and Grohmann, 2011). Pausing is believed to play a fundamental role in gene expression (Landick, 2009), for example by enabling the recruitment of transcription factors and by enabling RNA folding and attenuation mechanisms to occur. Pausing also means that elongation by RNAP is inherently inefficient and discontinuous. In order to overcome this, elongation requires processivity and transcript cleavage factors. RNAP backtracking results in the 3′ end of the transcript protruding into the NTP-binding site, preventing the addition of further nucleotides (Komissarova and Kashlev, 1997). To overcome this the RNAP must cleave the 3′ end of the transcript resulting in a new 3′ terminus at the active site. While RNAP can carry out this exonucleic cleavage by itself, it

is aided by the cleavage factors. These are called Gre factors in bacteria and TFS in archaea (TFIIS in eukaryotes). The bacterial and archaeal/eukaryotic cleavage factors share no sequence or structural identity. However they function in the same manner. GreB in bacteria comprises an N-terminal and a C-terminal domain. The N-terminal domain of GreB is a coiled-coil that binds to the RNAP secondary channel and extends 45 Å into the active site. Two absolutely conserved acidic residues (one aspartate, one glutamate) play a critical role through interactions with the two Mg^{2+} ions, stimulating transcript cleavage (Opalka et al., 2003). In eukaryotes TFIIS contains three domains (domain I, domain II, and domain III) and an inter-domain linker between domains II and III. Domain I is poorly conserved and not required for TFIIS activity (Nakanishi et al., 1995). Domain II is α helical in structure. Along with the inter-domain linker, it is required for binding to RNAP subunits RPB1 and RPB9. Domain III is folded into a zinc-ribbon. Like GreB, it extends into the RNAP secondary channel where two acidic residues at the end of a hairpin (one aspartate, one glutamate) are in close proximity to Mg-1 (Kettenberger et al., 2003), resulting in promotion of RNAP nuclease activity (Figure 1.10). Whole-genome sequencing revealed that TFIIS also functions on RNAP III-expressed genes (Ghavi-Helm et al., 2008). The bacterial Gre factors are non-essential, since knocking them out is not lethal to cells (Stepanova et al., 2007). In eukaryotes TFIIS is also non-essential. In yeast, TFIIS knockouts are viable but display increased sensitivity to oxidising substances (Koyama et al., 2003). This suggests that intrinsic transcript cleavage by RNAP is important for viability (Sigurdsson et al., 2010).

1.6 Transcription Termination

1.6.1 Termination in Bacteria

Transcription termination can occur by two mechanisms in bacteria: intrinsic and factor-dependent. Intrinsic termination is mediated by RNA secondary structure and is encoded by a GC-rich hairpin, followed by a polyU stretch. It is believed that the polyU:polyA RNA:DNA hybrid, which has fewer hydrogen bonds than hybrids containing GC regions, facilitates the disengagement of RNAP from the ternary complex. Factor-dependent termination requires a hexameric helicase called rho.

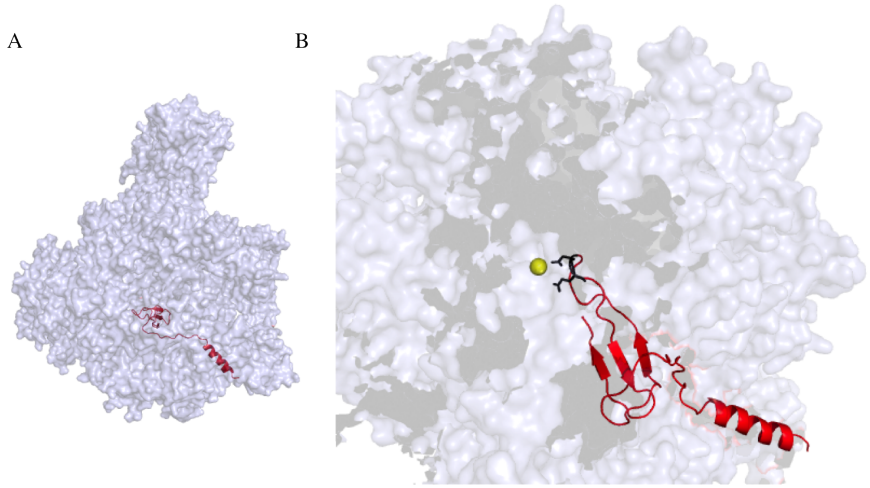


Figure 1.10: TFIIS bound to RNAP from *S. cerevisiae* (PDB 1Y1V). TFIIS binds to the secondary channel of RNAP, where it extends into the active site. Two conserved acidic residues (black) coordinate Mg-1, stimulating transcript cleavage. A) Complete RNAP structure shown. B) Close-up showing coordination of the magnesium ion by TFIIS. RNAP subunits RPB1, RPB2, and RPB9 only are shown.

There is ongoing debate regarding the mechanism by which rho acts to stimulate efficient termination. Classical models postulate the rho hexamer translocating along the emerging transcript in a 3′-5′ direction until it collides with the RNAP pushing it away from the RNA. Alternatively ATP hydrolysis results in the RNA being pulled out of the active site. More recently an allosteric mechanism has been suggested. In this model rho can bind to RNAP in the absence of RNA. UV crosslinking demonstrated close proximity between rho and the large catalytic β and β' subunits. The ratio of rho-β and rho-β' crosslinked adducts changes depending on whether the RNAP is paused at a rho utilisation (*rut*) site. The crosslinking on the β' subunit was then mapped to the trigger loop. Since this is a highly mobile element, it was suggested that rho binding might affect conformational changes in the trigger loop. Mutants of the trigger loop were then shown to affect the efficiency of rho termination and antibiotics that stabilise the closed trigger loop conformation prevented rho termination occurring (Epshtein et al., 2010). Other studies, however, present conflicting evidence, suggesting that rho and RNAP form part of the same complex only in elongation complexes where the nascent RNA is long and contains a *rut* site (Kalyani et al., 2011).

1.6.2 Termination in Eukaryotes

In eukaryotes transcription termination requires a 5′-3′ exoribonuclease called Rat1/Xrn2 which is recruited to RNAPII via the poly(A) stretch at the 3′ end of the transcript and also to the C-terminal tail of RNAPII (Kim et al., 2004; West et al., 2004). After transcribing the poly(A) tail, RNAPII pauses and recruits proteins such as CPSF which process the 3′ end of the transcript. The 3′ processing enzymes are thought to create a free 5′ end for the Rat1/Xrn2 which can degrade the RNA until it reaches the RNAPII, disengaging it from the DNA (Richard and Manley, 2009).

1.6.3 Termination in Archaea

In archaea termination occurs at a polyU stretch within the transcript (poly dT in the template DNA) and the Rpo4/7 stalk increases termination efficiency *in vitro* (Hirtreiter et al., 2010a). Termination still occurs when the terminator sequences are not predicted to form hairpins, unlike in bacteria (Santangelo et al., 2006). No factor-dependent termination has been observed in archaea. However, the fact that bacteria and eukaryotes both have factor-dependent termination makes it likely that archaea do too. One candidate which may function in termination is Mj0669. Mj0669 is a putative DEAD box RNA helicase which contains motifs characteristic of helicases such as the Walker A and Walker B sequences (Story et al., 2001). A crystal structure has been solved but no functional analyses have been carried out.

1.7 The Universally Conserved Elongation Factors Spt4/5 and NusG

Spt4/5 is the only universally conserved RNAP-associated transcription factor (Werner and Grohmann, 2011) and as such is one of the most ancient proteins involved in transcription regulation (Harris et al., 2003). In bacteria NusG (the Spt5 homologue) consists of two domains: a NusG N-terminal domain (NGN) and a C-terminal Kyrpides-Ouzounis-Woese domain (KOW) (Kyrpides et al., 1996) (Figure 1.11A). In *E. coli* NusG is an essential protein. However, in

Bacillus subtilis and *Staphylococcus aureus* there is no loss of viability when NusG is deleted (Ingham et al., 1999; Xia et al., 1999). Archaea and eukaryotes retain the core structure found in bacteria, with a number of additions. Archaeal and eukaryotic Spt5 is bound at the NGN domain by a smaller protein called Spt4 (Hirtreiter et al., 2010a; Ivanov et al., 2000; Wenzel et al., 2010) (Figure 1.11B). Unlike Spt5, Spt4 is not essential for cell viability (Malone et al., 1993). No Spt4 homologue has been discovered in bacteria (Guo et al., 2008). Eukaryotic Spt5 also contains multiple KOW domains and C-terminal heptad repeats which can be phosphorylated (Ivanov et al., 2000) (Figure 1.11C). In humans the heptad repeats function in an analogous manner to the CTD of RNAPII, with phosphorylation of Thr-4 by P-TEFb being a requirement for the activation of the elongation-enhancing properties of Spt5 (Yamada et al., 2006).

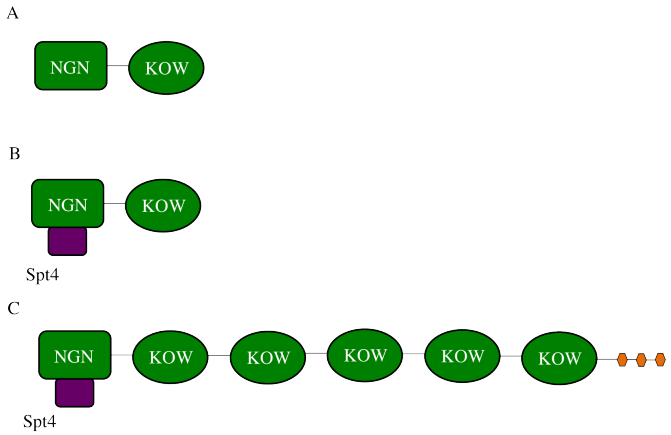


Figure 1.11: Schematic representation of NusG and Spt4/5 across the three domains of life. A) Bacterial NusG contains an N-terminal NGN domain and a C-terminal KOW domain. B) In archaea, the NGN domain of Spt5 is bound by Spt4 (purple). C) Eukaryotic Spt4/5 contains multiple KOW domains and heptad repeats (orange) at its C terminus.

The NGN domain of Spt5 from *M. jannaschii* consists of a four-stranded antiparallel β sheet, surrounded by three α helices. The KOW domain of Spt5 is comprised of five β strands which fold to form a compact barrel. The two domains do not interact with each other (Burmann et al., 2011; Mooney et al., 2009b) and are connected by a flexible linker (Figure 1.12). Spt4 is composed of a four-stranded antiparallel β sheet and interacts with the NGN domain of Spt5 via a hydrophobic interface forming a continuous eight-stranded β sheet. Four conserved cysteines in Spt4 coordinate a zinc atom (Hirtreiter et al., 2010a). The NGN domain of Spt5 does not undergo any structural reorganisations upon binding Spt4 (Zhou et al., 2009).

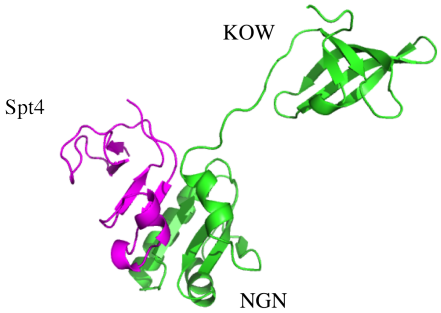


Figure 1.12: Homology model of *M. jannaschii* Spt4/5. Spt5 is in green. Spt4 is stably associated with the Spt5 NGN domain.

1.7.1 Spt4/5 and NusG are Versatile Multifunctional Transcription Factors

The primary function of Spt4/5 is to increase the processivity of RNA polymerase, and it stimulates transcription elongation *in vitro* (Hirtreiter et al., 2010a). The NGN domain of Spt5 is necessary for elongation stimulation in archaea, but not sufficient. In addition it requires either the KOW domain or Spt4 (Hirtreiter et al., 2010a). Site-directed mutagenesis studies demonstrated that Spt4/5 binds to the coiled-coil of the RNAP clamp via a small hydrophobic patch on the NGN domain (Hirtreiter et al., 2010a). The clamp is a regulatory 'hotspot' and is bound by a number of factors that modulate transcription, including the initiation factor TF(II)E (Grohmann et al., 2011), the $\sigma 2$ domain in bacteria (Sevostyanova et al., 2008) and the TFIIB linker region in eukaryotes (Kostrewa et al., 2009). NusG proteins also make contacts with the β gate loop of RNAP which results in the enclosing of the DNA-binding channel (Sevostyanova et al., 2011) and prevents nucleic acid escape from the elongating RNAP. Vice versa Spt4/5 binding to non-transcribing RNAP inhibits DNA loading and non-specific initiation (Grohmann et al., 2011). This may not hold true for eukaryotic Spt4/5, where *in vitro* data suggests that Spt4/5 can only efficiently bind elongation complexes that have already transcribed at least 30 nucleotides of RNA (Cheng and Price, 2008). However, ChIP analysis indicates that Spt4/5 is a component of the elongation complex immediately after promoter escape suggesting that it can bind to RNAP in the absence of a long transcript (Mayer et al., 2010).

1.7.2 Structure of RNAP-Spt4/5 Complexes

A cryoEM structure showed that when Spt4/5 is bound to RNAP, it encloses the active site (Klein et al., 2011). A high-resolution crystal structure of Spt4/5 NGN in complex with a recombinant RNAP clamp domain has also been obtained. This was then fitted to RNAP structures, confirming the conclusions drawn from the cryoEM data and providing much more detailed information on the specific interactions between Spt4/5 and RNAP (Martinez-Rucobo et al., 2011) (Figure 1.13).

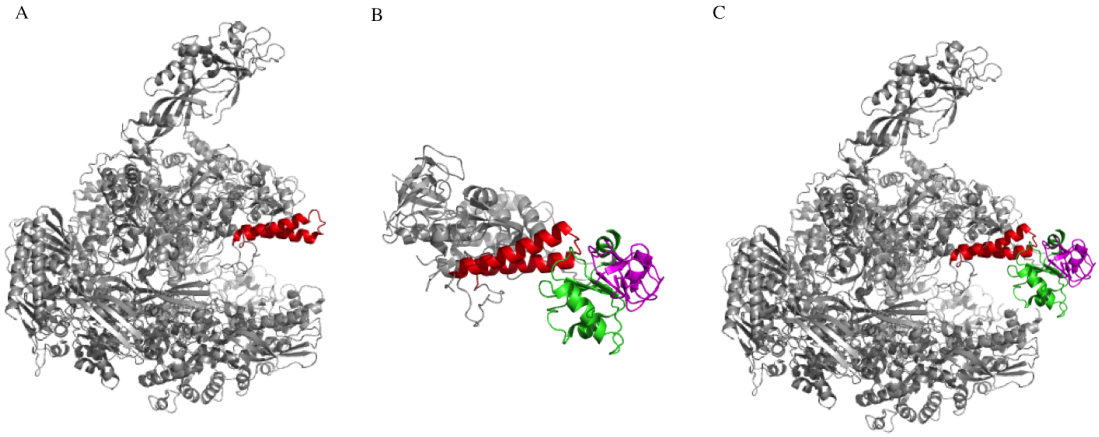


Figure 1.13: Spt4/5-RNAP complex structures. A) Crystal structure of *S. shibate* RNAP (PDB 2Y0S). The clamp coiled-coil helices are indicated in red. B) Crystal structure of *P. furiosus* Spt4/5 bound to recombinant RNAP clamp (PDB 3QQC). Spt5 NGN: green, Spt4: purple. C) Fitting of the *P. furiosus* Spt4/5 recombinant clamp structure to the *S. shibate* RNAP structure.

In addition to the structural model, an allosteric mechanism of NusG-like proteins has also been proposed (Svetlov et al., 2007). The coiled-coil to which the NGN domain of Spt5 binds is part of the RNAP clamp and is 75 Å from the active centre. At the base of the clamp is the bridge helix and trigger loop, which are mobile elements that are important for translocation and substrate selection respectively. Therefore Spt4/5 binding to the coiled-coil of the clamp may induce a conformational change within the clamp, the effects of which could be amplified in the trigger loop and active centre, thereby increasing the processivity of RNAP.

1.7.3 Molecular Mechanisms of Spt4/5 During Promoter Escape

The transcription initiation factor TFE binds to the same region of RNAP as Spt4/5. Single molecule FRET based NPS analysis showed that the winged helix domain of TFE, like the NGN domain of Spt5, interacts with the tip of the RNAP coiled-coil, while the zinc ribbon domain of TFE interacts with the clamp and Rpo4/7 stalk of RNAP. Mutagenesis of the RNAP confirmed these binding sites. It was then shown that TFE and Spt4/5 compete for the binding of this site. The addition of Spt4/5 to the pre-initiation complex inhibits promoter-directed transcription initiation. However in order to displace TFE in the pre-initiation complex Spt4/5 had to be at a 50 times higher concentration than TFE. This shows that in the initiation complex RNAP has a higher affinity for TFE than for Spt4/5. However, once promoter escape has occurred RNAP has a greater affinity for Spt4/5 over TFE and 'swaps' the transcription factors in favour of Spt4/5 (Grohmann et al., 2011). In bacteria the initiation factor σ^{70} competes with the NusG paralogue RfaH for binding to the RNAP clamp in an analogous manner (Sevostyanova et al., 2008).

1.7.4 Molecular Mechanisms of NusG During the Coupling of Transcription and Translation

In prokaryotic cells there is no compartmentalisation and the processes of transcription and translation are not separated by a nuclear membrane, enabling them to be coupled in bacteria and archaea (French et al., 2007). This allows the transcriptional output to be adjusted for the translational requirements. This increases energy efficiency which is advantageous for single-celled species that are growing under stress conditions and that are in constant competition with each other for resources. As soon as the ribosome-binding site on the transcript emerges, the ribosome can initiate translation while the RNA is still being transcribed. There have been reports of transcription-translation co-regulation for over 30 years. In an operon, distal genes are often less frequently transcribed. This is because during gene expression the distance between RNAP and the ribosome increases, resulting in increased termination by rho (Richardson, 1991). This phenomenon is called polarity.

Attenuation

Attenuation is a mechanism in which the ribosome regulates transcription. The *trp* operon contains the genes involved in the synthesis of tryptophan (Yanofsky, 1981). During transcription, the *trp* leader RNA can fold in two conformations. The first is a termination hairpin that prevents expression of the operon. The second is an antiterminator structure that allows the RNAP to express the genes within the operon. The levels of tryptophan present in the cell regulate its further expression. The operon leader sequence, *trpL*, is about 130 nucleotides long and contains four sequences that are complementary to each other. The first of these sequences contains two adjacent codons for tryptophan. When tryptophan is abundant in the cell there is enough charged $tRNA^{Trp}$ for the ribosome to read through the two Trp codons rapidly. The ribosome occludes sequences 1 and 2. This results in the termination hairpin being formed between sequences 3 and 4 and the RNAP will not transcribe the operon (Figure 1.14A). If the levels of tryptophan in the cell are low then the ribosome will stall at the two tryptophan codons as there will not be enough charged $tRNA^{Trp}$. When the ribosome stalls here it occludes sequence 1 only, enabling the antitermination hairpin to form between sequences 2 and 3 and thereby preventing formation of the termination hairpin. Therefore RNAP is able to transcribe the genes of the operon (Figure 1.14B). Thus the rate of translation of the leader sequence determines the secondary structure of the RNA and controls transcription of the operon. This mechanism is not exclusive to the *trp* operon. The *his* operon, which contains the genes required for histidine biosynthesis, contains seven adjacent His codons. In a similar fashion to the *trp* operon, when histidine is abundant, the ribosome reads through the histidine codons and a terminator structure forms preventing expression of the operon. When histidine is scarce the ribosome pauses at one of the seven histidine codons, occluding the terminator from forming and enabling expression of the operon (Chan and Landick, 1993, 1989).

More recently, it has become apparent that not only are transcription and translation co-regulated but that they are also directly coupled through their respective machinaries, RNAP and the ribosome. An NMR study by Burmann et al. analysed the interaction between the ribosomal protein S10 and NusG (Burmann et al., 2010b). HSQC experiments demonstrated that the KOW domain of NusG binds to S10 mainly through hydrophobic interactions (Figure 1.15). Titration experiments indicate a K_d of 50 μM for this interaction. S10 is a component of the

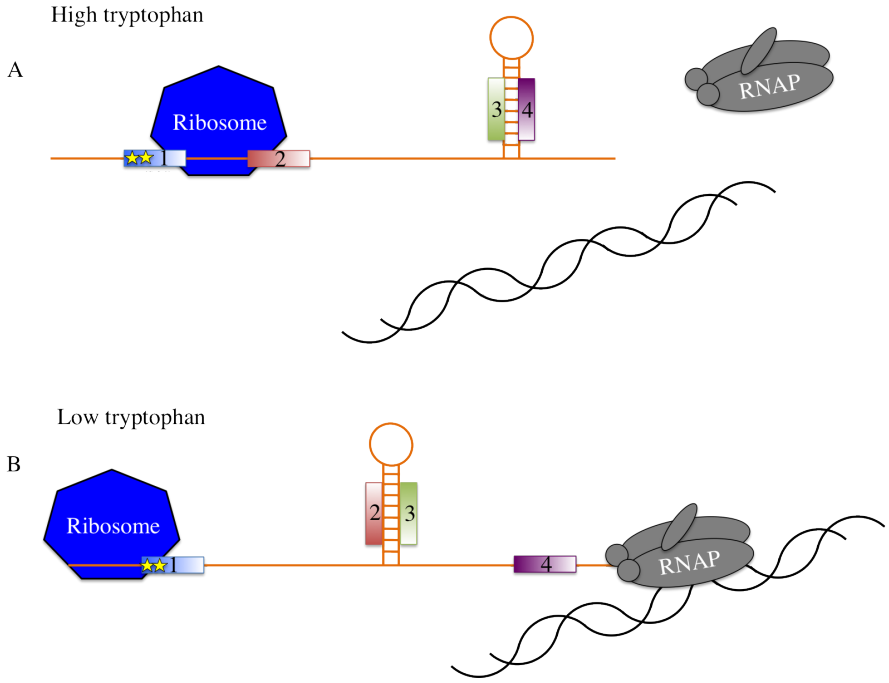


Figure 1.14: Mechanism of attenuation on the tryptophan operon. A) When tryptophan is abundant, the ribosome is able to readily translate through the two adjacent tryptophan codons (yellow stars) of sequence 1, and it occludes sequences 1 and 2. Therefore a terminator hairpin forms between sequences 3 and 4, preventing further transcription of the operon by RNAP. B) When tryptophan is low, the ribosome stalls at the two adjacent tryptophan codons of sequence 1, and it occludes sequence 1 only. Therefore a hairpin forms between sequences 2 and 3, and no terminator is formed. This enables expression of the operon.

small ribosomal subunit and has also been identified as an antitermination factor (Friedman et al., 1981).

The ribosome binding face of S10 is still accessible when it is in complex with NusG and the authors propose a model whereby NusG acts to bridge the translating ribosome to the elongating RNAP. Taking an alternative approach, a study by Proshkin et al. uses an *in vivo* system to monitor the rates of translation elongation and transcription elongation (Proshkin et al., 2010). Using antibiotics to specifically modulate ribosome translocation they demonstrated that a decrease in the rate of translation is accompanied by a decrease in the rate of transcription. Additionally mutants with slow ribosomes also showed a decreased rate of transcription. The use of antibiotics to partially rescue this mutant resulted in a recovery of transcription rates. In summary, they showed that the rates of transcription and translation are matched so that for

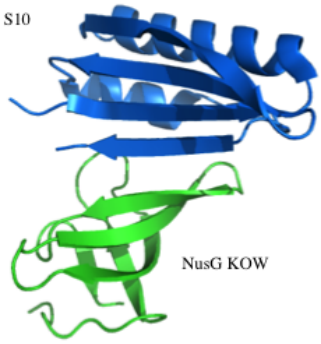


Figure 1.15: NMR structure of the NusE-NusG complex. NusE (S10) is in blue, NusG KOW domain is in green. PDB 2KVQ.

every amino acid incorporated into the nascent chain, three nucleotides are added to the transcript (Proshkin et al., 2010). Since each codon consists of three nucleotides this provides very strong evidence that transcription and translation are coupled in *E. coli*. An *in vitro* system to study the coupling of transcription and translation has been developed for *E. coli* which utilises purified ribosomes and recombinant transcription/translation factors. This has the potential to explore the phenomenon of transcription-translation coupling in molecular detail. However, the role of NusG has not yet been analysed in this system (Castro-Roa and Zenkin, 2012).

The ribosome also helps prevent RNAP backtracking and facilitates the read-through of roadblocks (Proshkin et al., 2010). RNAP can get stuck in an unproductive backtracked state where the 3′ end of the transcript is no longer in the active site (Komissarova and Kashlev, 1997; Reeder and Hawley, 1996). This prevents the addition of further nucleotides. The ribosome 'pushes' RNAP forward and out of the backtracked state (Proshkin et al., 2010). This complements earlier data which showed that if RNAP is stuck in a backtracked state, a second transcribing RNAP behind it can push it out of this state with no cleavage of RNA required (Epshtein and Nudler, 2003; Epshtein et al., 2003).

S10 Interactions Within the Ribosome

Within the ribosome S10 makes multiple protein-protein, and protein-RNA contacts (Figure 1.16). It contacts the ribosomal proteins S3 and S14, and it contacts RNA helices H31, H34, H38, H39, H41, and H43 of the 16S rRNA. A long β hairpin protrudes deep into the ribosome (Wim-

berly et al., 2000). The structure of S10 within the ribosome (Figure 1.16D) initially appears dramatically different and much more elongated compared to that of free S10 (Figure 1.16E). However, the apparent differences are because the elongated ribosome-binding loop was deleted in the structure of free S10 (Luo et al., 2008). In both structures, the core globular domain of S10 contains two α helices. In ribosome-incorporated S10, the rest of the core appears not to adopt a canonical secondary structure. In free S10, this region comprises four β strands which form a continuous antiparallel sheet. However, the overall conformation of the S10 core is similar in both structures, suggesting that it does not undergo global refolding to convert it from a ribosomal protein into an antitermination factor (Luo et al., 2008). In both free S10 and ribosome-incorporated S10, the residues that interact with the NusG KOW domain are accessible (Figure 1.17).

1.7.5 NusG as a Suppressor of Foreign DNA

The NusG KOW domain interacts with the termination factor rho (Burmann et al., 2010b; Li et al., 1993) and stimulates transcription termination (Sullivan and Gottesman, 1992). One study, looking at the effects of rho on a whole-genome level, found that when rho is inhibited by the antibiotic bicyclomycin there is an increase in expression of non-coding regions and horizontally transferred genes from bacteriophage and other species. When these horizontally transferred genes are deleted to make a synthetic strain, MDS42, *E. coli* becomes much more resistant to bicyclomycin treatment. One of the deleted genes in this strain is the lambdoid bacteriophage gene *rac*, which inhibits cell division. If *rac* alone is deleted, bicyclomycin resistance also occurs, suggesting that rho acts to prevent expression of this lethal gene. In cells where *kil* is deleted, *nusG* and *nusA* can also be deleted, even though they are essential to the wild-type cell. This suggests that NusG and NusA act with rho to terminate transcription of the *kil* gene (Cardinale et al., 2008).

1.7.6 RfaH - A Bacterial NusG Paralogue

RfaH is a paralogue of NusG. Like NusG it increases the processivity of RNAP and reduces RNAP pausing at regions where backtracking occurs. However, while NusG is involved in the

Figure 1.16: S10 in the context of the ribosome. A) The 30S subunit of the *T. Thermophilus* ribosome. The S10 protein is in blue and the remaining protein and 16S rRNA is in grey. B) S10 makes protein-protein interactions with S3 (pink) and S14 (cyan) within the ribosome. C) S10 interacts with the 16S rRNA (orange). D) Ribosome-incorporated S10 structure. E) Free S10 without the ribosome-binding loop. A-D) PDB 1FJG (Carter et al., 2000), E) PDB 3D3B (Luo et al., 2008).

expression of most genes, RfaH acts only on a small set which contain a 12 nucleotide *ops* sequence. RfaH contains two domains that are connected by a flexible linker. The N-terminal domain is structurally similar to that in NusG. The C-terminal domain adopts an α helical structure, unlike NusG. The two domains of RfaH are associated via a hydrophobic interaction resulting in a closed inactive conformation. When RfaH binds the non-template strand at the *ops* site the two domains separate (Belogurov et al., 2007). The RNAP-binding surface of RfaH is exposed only after *ops*-dependent isomerisation has occurred. At the *ops* site RNAP pauses and RfaH is recruited and loaded, even in the presence of excess NusG (Belogurov et al., 2008). RfaH functions as an antiterminator, suppressing the action of rho. The antipausing effects at rho-termination sites are mediated solely through the N-terminal domain (Belogurov et al., 2008).

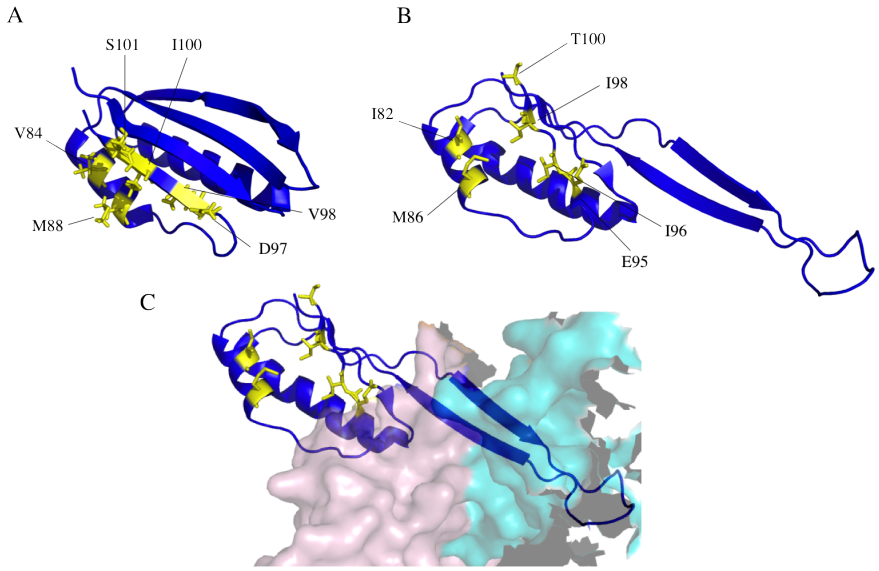


Figure 1.17: Availability of the NusG-binding face of S10. A) Structure of *E. coli* S10 (PDB 2KVQ). The residues identified as interacting with the NusG KOW domain are highlighted in yellow. B) Structure of S10 within the small ribosomal subunit from *T. thermophilus*. The corresponding residues that interact with NusG are highlighted in yellow (PDB 1FJG). C) Structure of S10 within the 30S ribosomal subunit. Ribosomal proteins S3 (pale pink) and S14 (cyan), which interact with S10, are shown as a space-filling model. The residues that are predicted to interact with NusG are accessible, suggesting that NusG can bind to ribosome-incorporated S10.

A recent NMR study has shed further light on the auto-regulation of RfaH. The solution structure of full length RfaH is in good agreement with the crystal structure showing that the N-terminal domain adopts a similar fold the NusG NGN domain, and that the C-terminal domain associates with the N-terminal domain and adopts an α helical structure (Figure 1.18A). However, when the C-terminal domain of RfaH is analysed in isolation it adopts a completely different secondary structure, comprising five β strands (Burmann et al., 2012). This is fitting with a previous bioinformatics analysis which suggested that the C-terminal domain was capable of folding into a β structure (Belogurov et al., 2008) (Figure 1.18B). The refolded RfaH C-terminal domain can interact with ribosomal subunit S10 and thereby couple transcription and translation (Burmann et al., 2012).

RfaH is the best characterised additional NusG-like protein but it is not unique in being a NusG paralogue. Multiple NusG paralogues which act in a sequence-specific manner have been identified. The bacterium *Bacteroides fragilis* has eight NusG paralogues, each responsible for

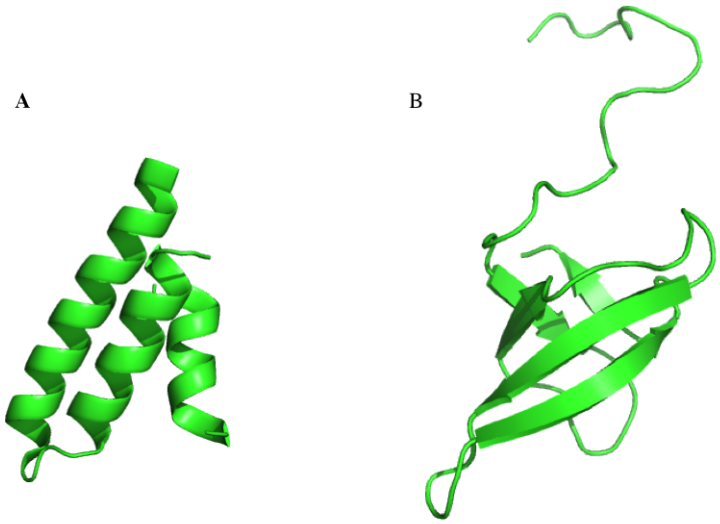


Figure 1.18: The RfaH C-terminal domain can adopt two folds. A) In free full-length RfaH (PDB 2OUG), the C-terminal domain of RfaH adopts an α helical structure. B) When the two RfaH domains separate, its C-terminal domain undergoes a complete refolding and adopts a β barrel structure (PDB 2LCL). This enables it to bind to the ribosome.

preventing premature termination during expression of different enzymes in a polysaccharide synthesis pathway (Chatzidaki-Livanis et al., 2009).

1.7.7 Functions of Spt4/5 in Eukaryotes

In eukaryotes Spt4/5 variants lacking one of the KOW domains are not able to bind RNAP as tightly, and the KOW domain alone exhibits low affinity binding to RNAP (Viktorovskaya et al., 2011). Crosslinking data suggests that in *S. cerevisiae*, the Spt5 KOW domains 4 and 5 interact with the RPB4/7 stalk (Li et al., 2014). This is in contrast to prokaryotes and archaea where the KOW domain has no RNAP-binding activity. The C-terminal repeats (CTR) that are present in eukaryotic Spt4/5 are also involved in the regulation of transcription. For example, when the CTR2 is removed there is an increase in basal transcription but also an increase in DRB-mediated inhibition (Ivanov et al., 2000). Eukaryotic Spt4/5 also binds to RNAPI where it can mediate positive and negative effects on transcription (Viktorovskaya et al., 2011).

Promoter-proximal Stalling

In metazoans (animals) Spt4/5 is involved in transcription regulation by promotor-proximal stalling (Figure 1.19). During this process, the initiation of transcription occurs before gene activation and RNAPII is poised near the promoter in a phosphorylation state associated with initiation rather than elongation. Promoter-proximal stalling was initially described through the use of DNA-protein cross-linking experiments (Rougvie et al., 1988). It was shown that RNAP is bound about 40 nucleotides from the promoter region of heat-shock genes such as *hsp70* even in the absence of heat-induction. Furthermore, by carrying out nuclear run-on assays it was determined that, not only were these RNAPs bound, they were transcriptionally active. It was subsequently shown that promoter-proximal stalling may occur, not only on inducible genes, but also on those which are constitutively expressed (Rougvie and Lis, 1990). The development of ultra-high-throughput sequencing enabled a genome-wide study of stalled genes, suggesting that around 30% of human genes undergo promoter-proximal stalling (Core et al., 2008).

The human homologue of Spt4/5 is called DSIF, an abbreviation of DRB sensitivity inducing factor. DRB is a kinase inhibitor that prevents phosphorylation of transcription factors, release of stalled complexes and thus transcription. DSIF, in conjunction with negative elongation factor (NELF), is part of the stalled RNAPII complex. Association of NELF to the complex requires DISF (Missra and Gilmour, 2010). In order to become elongation competent NELF, DSIF and RNAP are phosphorylated by a protein called P-TEFb. Although it is not known why the stalling occurs around 40 nucleotides downstream of the promoter, one hypothesis suggests that it is connected to the fact that the RPB4/7 complex interacts with around 40 nucleotides of transcribed RNA (Újvári and Luse, 2005).

It is widely thought that promoter-proximal stalling is important for enabling a rapid induction of gene expression in response to external stimuli, such as TNF-α expression in response to an antigen (Adelman et al., 2009). However it may have other roles in promoting gene expression. For example, when NELF levels are dramatically reduced by RNAi, a nucleosome appears over the promoter preventing RNAP access in *TepII* (Gilchrist et al., 2008).

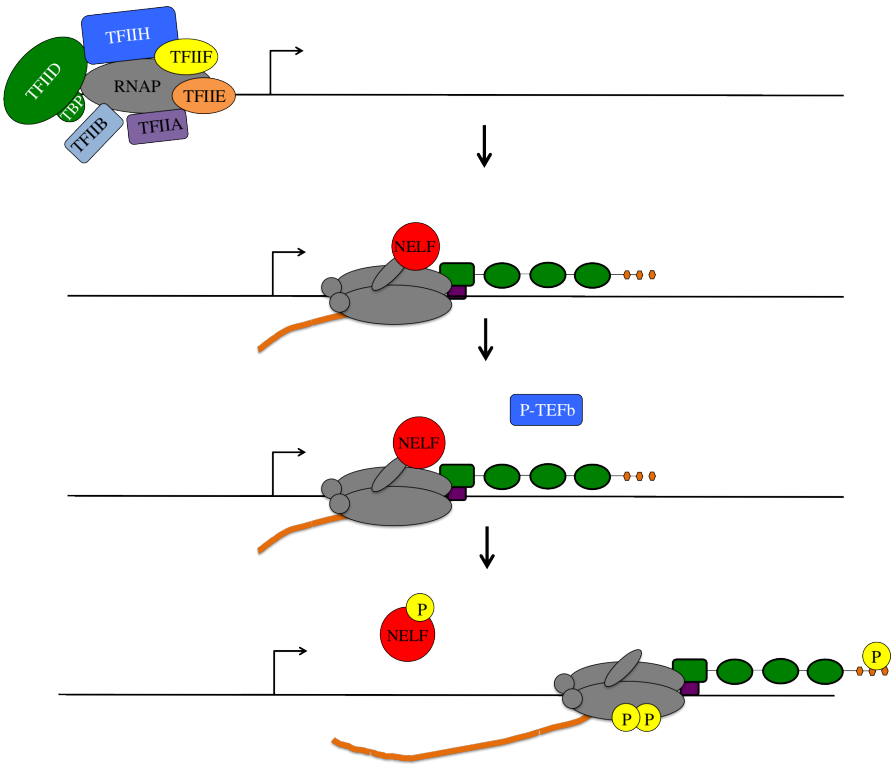


Figure 1.19: Mechanism of promoter-proximal stalling. After initiation RNAP transcribes approximately 40 nucleotides of RNA. Association with DSIF and NELF (red) results in the stalling of RNAP slightly downstream of the promoter. The complex remains stalled until RNAP, NELF and DSIF are phosphorylated by P-TEFb (blue). After phosphorylation RNAP continues transcribing the gene.

Nucleosome Remodelling and RNA Processing

The KOW domains of eukaryotic Spt5 enable the recruitment of multiple transcription factors with a range of functions. Spt5 copurifies in complexes containing Spt6 (Lindstrom et al., 2003), a protein that has multiple roles including the restoration of nucleosomes on highly transcribed genes (Ivanovska et al., 2011). Human Spt4/5 interacts with the protein Paf1, which is itself important for generating histone modifications associated with transcription such as H2B K123 monoubiquitination and H3 K4, K36 and K79 trimethylation (Chen et al., 2009). Spt4/5 is also genetically linked with *CET1* and *CEG1* (Lindstrom et al., 2003), which between them form the 5′ RNA capping complex that is required to add a 5′ guanine-N7 cap in eukaryotes. Furthermore, Spt4/5 has also been observed to copurify with the RNA capping enzyme, suggesting a function for Spt4/5 in pre-mRNA processing.

Function	Bacteria	Archaea	Eukaryotes
Increase RNAP processivity	✔	✔	✔
Coupling of transcription and translation	✔	?	
Factor-dependent termination	✔	?	
Antitermination	✔	?	
Silencing of foreign DNA	✔	?	
Promoter-proximal stalling		?	✔
RNA processing		?	✔
Chromatin remodelling		?	✔

Table 1.2: Summary of known NusG and Spt4/5 functions across the three domains of life.

1.8 Transcription and Translation are Coupled in Archaea

There is strong evidence that transcription and translation are coupled in the euryarchaeota *Thermococcus kodakaraensis* (Santangelo et al., 2008). *T. kodakaraensis* is naturally competent. Mutations can be introduced via homologous recombination. A gene expression reporter system was developed in *T. kodakaraensis* to enable the analysis of transcription and translation *in vivo*. Transcription was monitored by microarray hybridisation and translation was monitored by β-glycosidase activity. A nonsense codon was introduced towards the $5'$ end of the reporter operon to stop translation. This also resulted in greatly reduced transcription of the operon. This demonstrates that there is polarity in archaeal operons and that transcription and translation are coupled. When the ribosome was stopped, transcription of the operon decreased (Santangelo et al., 2008). The mechanisms behind the coupling remain unknown. However, given the conservation of NusG proteins, there is a sound rationale for studying the possible role of Spt4/5 in the coupling of transcription and translation in archaea.

1.9 Structure and Function of NusA

In *E. coli*, *nusA* is an essential gene, and homologues have been identified in all sequenced bacterial genomes. Bacterial NusA comprises an RNAP-binding N-terminal domain, an S1 domain and two RNA-binding KH domains. Evidence suggests that NusA probably functions as a monomer (Gopal et al., 2001). Its N-terminal domain binds to the β flap of RNAP (Mah

et al., 1999) and, when acting in isolation, NusA promotes RNAP pausing at RNA hairpins and stimulates termination (Schmidt and Chamberlin, 1987; Toulokhonov et al., 2001). In *E. coli* and *M. tuberculosis*, NusA is associated with RNAP throughout the entire genome, including promoter-proximal peaks (Mooney et al., 2009a; Uplekar et al., 2013). This suggests that NusA is bound to both actively transcribing and stalled RNAP complexes.

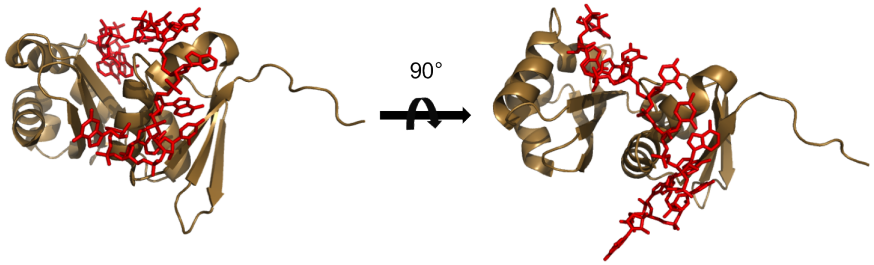


Figure 1.20: *M. tuberculosis* NusA KH domains bound to RNA (PDB 2ASB). The RNA binds across both KH domains in a continuous, elongated conformation.

1.9.1 Regulation of Gene Expression by Antitermination of Transcription

Whilst NusA acts to promote pausing on its own, it is also a component of the antitermination complex. Antitermination complexes enable RNAP to enter a pause-resistant state and to transcribe through termination sequences. Antitermination was first described in bacteriophage lambda genes. A phage-encoded protein called N enables RNAP to read through intrinsic termination sequences of the early-stage phage genes and transcribe the middle-stage genes. While the N protein can bind RNAP by itself, the complex is unstable. The addition of host Nus (N-utilisation substance) factors NusA, NusG, NusE (S10), and NusB results in the formation a stable antitermination complex with stronger antitermination properties (Dodd et al., 2005; Gusarov et al., 2001; Mason and Greenblatt, 1991) (Figure 1.21). The Nus factors were first identified by searching for *E. coli* mutants that were unable to support N-mediated antitermination (Friedman et al., 1981, 1976). The proteins were subsequently isolated. The NusA protein was first purified by using λN to pull down proteins from *E. coli* cell extract and searching for a protein that had the correct predicted mass (Greenblatt and Li, 1981). It was later shown that

the addition of NusA, B, E, and G enable RNAP to read through terminator sequences *in vitro* (DeVito and Das, 1994).

The formation of an antitermination complex also requires a conserved region of RNA called the N-utilisation site (*Nut*). The *Nut* site is found towards the 5′ end of a transcript and is composed of a sequence called BoxA and an RNA hairpin called BoxB, which are separated by a spacer region. BoxB is bound by the N-terminal domain of N (Legault et al., 1998; Schärpf et al., 2000). NusA then binds directly to N with its C-terminal domain (Mah et al., 1999) and to the RNA spacer between BoxA and BoxB with its KH domains. BoxA is bound by a dimer of S10 and NusB (Lüttgen et al., 2002). Mutations within BoxA prevent this association and result in impaired antitermination (Heinrich et al., 1995). NusB forms a stable complex with NusE (S10) (Mason et al., 1992) and the two proteins coelute in affinity purifications. Upon forming a dimer with S10, NusB binds RNA with 10-fold higher affinity than NusB alone (Burmann et al., 2010a). The antitermination function of S10 is mutually exclusive from its ribosomal function. When S10 is within the 30S subcomplex its NusB binding face is occluded (Luo et al., 2008). This is supported by the fact that when the ribosome binding loop of S10 is substituted with a serine, it is still able to mediate antitermination (Luo et al., 2008).

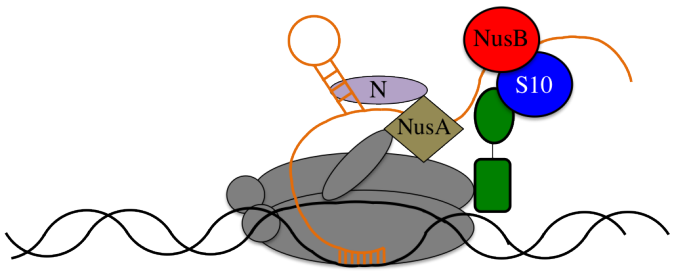


Figure 1.21: Schematic of an antitermination complex. A Phage protein called N (mauve) binds to an RNA hairpin called BoxB towards the 5′ end of the transcript. N is bound by NusA (beige), which is itself bound to RNAP (grey). In addition NusB (red) loads NusE/S10 (blue) on to the BoxA RNA. The KOW domain of NusG (green) is also bound to S10.

Phage λ Antitermination Factor Q

Bacteriophage λ have another antitermination factor called Q, which functions to enable the late phage genes to be expressed. After initiation at the late promotor (P_R), RNAP pauses about

16 nucleotides downstream from the transcription start site. This pausing is necessary to enable DNA-bound Q to also bind to RNAP (Ring and Roberts, 1994). Q initially interacts with the σ factor (Nickels et al., 2002), and subsequently displaces it to bind to the β flap of RNAP (Deighan et al., 2008), becoming stably incorporated into the elongation complex. The precise mechanism of Q action remains unclear but it is known that NusA stimulates its activity (Yarnell and Roberts, 1992). It is likely that Q and NusA bind to the RNA transcript, preventing the formation of hairpin terminators and the action of rho (Santangelo and Artsimovitch, 2011).

1.9.2 Antitermination on Ribosomal Operons

The ribosomal operons account for up to 60% of cellular transcription (Bremer et al., 1996) during exponential growth. Robust expression of the rRNA genes is therefore extremely important. There are 7 ribosomal operons in *E. coli*. Each operon contains the 16S rRNA gene, a tRNA gene, the 23S rRNA gene and the 5S rRNA gene. The rRNA genes are not translated, thereby potentially exposing RNAP to the termination factor rho. However, antitermination also occurs on ribosomal RNA operons to prevent this. Antitermination complex formation on ribosomal operons also requires the proteins NusA, NusB, NusE and NusG. Ribosomal operons contain a BoxA-like region at their leader that can be bound by a complex of NusB and S10 (Nodwell and Greenblatt, 1993). NusA also binds close to this region in the *rrn* operon (Vogel and Jensen, 1997). In addition several other ribosomal proteins have been implicated (Condon et al., 1995), for example S4 binds to RNAP and has been shown to have antitermination properties similar to NusA (Torres et al., 2001). NusA mutant strains show increased polarity on the rRNA operons with two-fold fewer RNAPs transcribing the downstream 23S rRNA compared to wild-type *E. coli* cells (Quan et al., 2005). This demonstrates the importance of NusA in antitermination of the ribosomal operons. Unlike λN antitermination which allows read through of rho-dependent and rho-independent terminators, *rrn* antitermination functions only on rho-dependent termination sites (Albrechtsen et al., 1990). No N-like protein has been implicated in antitermination of ribosomal operons and its precise mechanism and regulation is poorly understood.

1.9.3 NusA in rRNA Processing

In *E. coli*, NusA is also required for processing of the pre-16S rRNA. The rRNA folds and is processed as it is being transcribed. The leader and downstream regions of the 16S gene are 1700 nucleotides apart and are complementary in sequence. They can therefore hybridise to form a dsRNA region with the 16S rRNA forming a loop at the end. A mismatch within the complementary double-stranded region is a target for cleavage by RNaseIII (Young and Steitz, 1978). RNaseIII processing is normally very rapid and occurs as soon as the two complementary regions of RNA have annealed. Mutations in NusA result in impaired biogenesis of the 30S subunit due to a buildup of 21S RNA, a precursor that is normally rapidly processed into the full 30S. This occurs because the rRNA is cleaved by RNaseIII in a misfolded conformation and suggests that NusA is required for the correct folding of the 16S RNA. When RNaseIII is also deleted, normal ribosome biogenesis resumes (Bubunenko et al., 2013). In the absence of RNaseIII, processing occurs by other RNases at a much slower rate, giving more time for the RNA to fold correctly. The role of NusA can be rationalised on a structural level. NusA tethers the rRNA 5′ end of the RNA by binding to the leader BoxA sequence, and being bound to RNAP itself. The 3′ end of the transcript is also constrained as it is bound within the active site. Thus the RNA is constrained at both ends whilst it is being transcribed, which facilitates loop formation and correct folding of the transcript (Figure 1.22).

1.9.4 Archaeal NusA

Little is known about the role of NusA in archaea. Sequence analysis indicates that NusA is much smaller in archaea than in bacteria. *M. jannaschii* NusA contains 183 residues compared to *M. tuberculosis* NusA which contains 347 residues. The crystal structure of *A. pernix* (archaeal) NusA demonstrated that it contains two KH domains (Shibata et al., 2007). While structurally similar to the KH domains from bacterial NusA there is little sequence identity, although BLAST analysis indicates genuine homology. Archaeal NusA does not retain the N-terminal RNAP-binding domain or the S1 domain. It is therefore not known whether archaeal NusA is able to promote RNAP pausing, act as a component of an antitermination complex, or even bind to RNAP. Of a small selection of RNA targets it was found to bind the 23S rRNA terminator sequence with very high affinity ($K_d = 5.6 \times 10^{-8}$ M) and to have a preference for

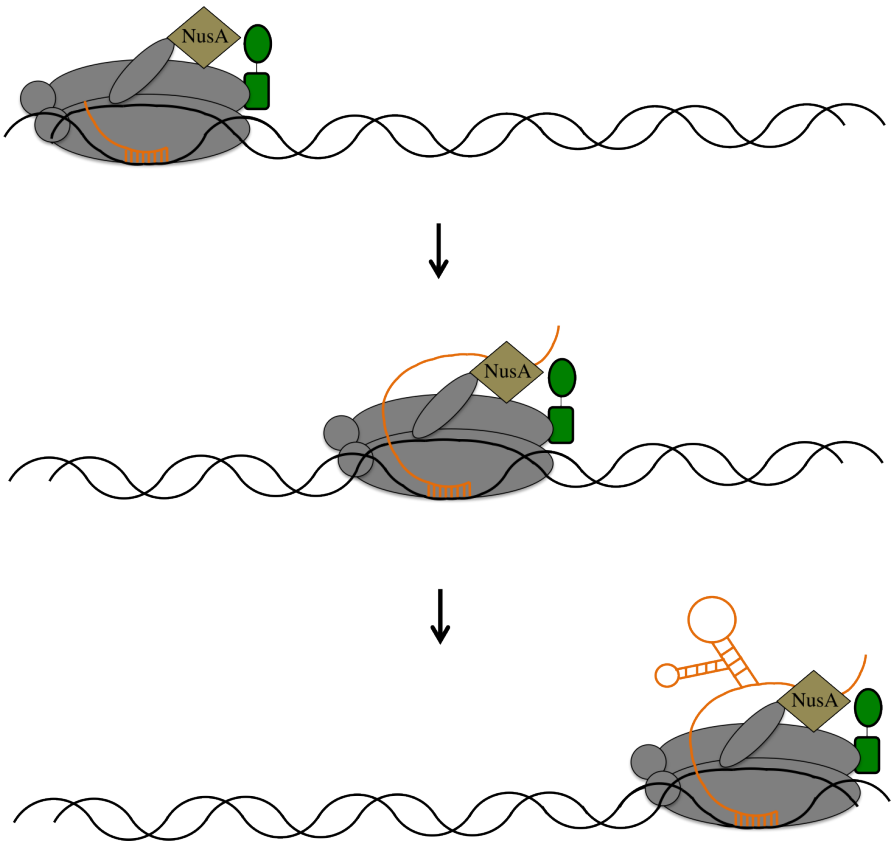


Figure 1.22: The role of NusA in ribosome biogenesis. NusA binds to the RNA transcript near the 5′ end, and to RNAP. The 3′ end of the transcript is within the active site. The nascent RNA is therefore constrained at both ends and forms a loop as it is transcribed, which facilitates RNA folding and therefore ribosome biogenesis.

pyrimidine-rich sequences. However, this does not provide any information about the *in vivo* target of archaeal NusA (Shibata et al., 2007).

1.9.5 Antitermination in Archaea?

There is very little data concerning antitermination in archaea. There is no nuclear membrane, so transcription and translation of a gene can occur at the same time (French et al., 2007). Archaeal genes are often organised and expressed in operons, similar to bacteria. Furthermore it has been shown *in vivo* that when translation of an operon is paused by introduction of a non-sense codon, transcription of downstream genes decreases (Santangelo et al., 2008). The fact that polarity exists in archaea raises the possibility that they have evolved mechanisms to

overcome it and to allow expression of distal genes in operons.

M. jannaschii has homologues of three of the Nus antitermination factors. These are NusA, NusE (S10), and Spt5 (NusG). *M. jannaschii* does not have a NusB homologue. However, NusB is not essential in *E. coli* (Taura et al., 1992). Therefore there is a sound rationale for studying NusA, Spt5, and S10 in archaea, with the aim of determining whether they are able to form a functional complex.

1.10 Summary

Transcription is a highly complex process in which a large array of proteins and nucleic acid motifs interact to carry out the genetic programme with astounding accuracy. RNAP is a multisubunit enzyme with a conserved structure and function across the three domains of life, with archaeal RNAP being similar to eukaryotic RNAPII. During elongation RNAP is prone to pausing, which on one hand can lead to transcription termination and on the other hand can be overcome in a regulated fashion through a plethora of transcription factors. Some of these factors are not conserved despite having similar roles. GreB in bacteria and TF(II)S in archaea/eukaryotes are a good example of this. Other factors are conserved, such as Spt4/5, which is the only known universally conserved RNAP-associated transcription factor. Its primary function is to increase the processivity of RNAP, which is supported by biochemical and structural data. This function is mediated by the NGN domain. However, NusG and Spt4/5 have additional functions, which are mediated by the KOW domain(s). In bacteria NusG couples transcription to translation by interacting with ribosomal protein S10 (NusE). It also supports factor-dependent transcription termination through its interaction with rho helicase and thereby reduces the expression of horizontally acquired genes. In eukaryotes Spt5 also binds to the RNAP clamp via its NGN domain. In addition eukaryotic Spt4/5 binds RNA-processing and nucleosome remodelling factors via its multiple KOW domains, and mediates promoter-proximal stalling. Thus the function of the NGN domain is conserved across the three domains of life whilst the KOW domain(s) have evolved separate functions. The KOW-mediated functions in archaea have not been studied in depth, and are poorly understood.

Bacteria	Archaea	Eukaryotes (RNAPII)	Function
NusG	Spt5	SPT5	**All:** Increases RNAP processivity **Bacteria:** Couples transcription and translation **Eukaryotes:** Promoter-proximal stalling
	Spt4	SPT4	**Archaea:** Stabilisation of Spt5 **Eukaryotes:** Unknown
NusA	NusA		**Bacteria:** Promotes termination by itself. Reduces termination efficiency as component of antitermination complex **Archaea:** Unknown
NusE (S10)	S10	S10	**All:** Ribosomal protein **Bacteria:** Binds NusG and component of antitermination complex
NusB			**Bacteria:** Forms dimer with S10. Component of antitermination complex
Gre factors	TFS	TFIIS	**All:** Stimulates transcript cleavage of stalled polymerases NB: Gre factors and TF(II)S are not homologous
Rho			Transcription termination

Table 1.3: Summary of transcription elongation factors across the three domains of life and their function.

NusA is an essential protein in *E. coli* which also has multiple functions. It contains an RNA-binding N-terminal domain, an S1 domain and two RNA-binding KH domains. In *E. coli*, NusA can bind to RNAP and stimulate transcription termination. However, it is also a component of the antitermination complex where it promotes antitermination. In phage λ-mediated antitermination, NusA forms an antitermination complex with the proteins NusG, NusB, NusE, and the phage protein N. Antitermination also occurs on ribosomal operons, where additional proteins such as S4 have been implicated. NusA has an important role in the processing of pre-ribosomal RNA and ribosome biogenesis. Archaeal NusA is smaller, containing two KH domains only. It is not known whether archaeal NusA acts as an antitermination factor or whether it binds to RNA with any degree of sequence specificity.

There is evidence that transcription and translation are coupled in archaea. Due to the evolutionary conservation of NusG proteins, there is a sound basis for studying the potential role of Spt4/5 in this process. In addition archaea have homologues of three of the four Nus factors. Antitermination complex formation in bacteria requires the presence of specific RNA sequences (BoxA and BoxB). There is therefore a sound rationale for analysing the RNA-binding of archaeal NusA to determine whether it functions as a component of an archaeal antitermination complex.

1.11 Aims of this Thesis

This thesis is concerned with the transcription factors Spt4/5 and NusA in *M. jannaschii*. The key aims and objectives are:

- To establish whether *M. jannaschii* Spt4/5 and NusA are found in complexes containing RNAP and ribosomes.
- To quantify and dissect the interaction between ribosomes and Spt4/5 *in vitro*.
- To determine whether *M. jannaschii* Spt5 exists in multiple conformations using EPR.
- To analyse and quantify the role of Spt4 in stabilising Spt5.
- To identify the regions of NusA that bind to RNA and assess their relative contributions to the interaction.

Chapter 2

Materials and Methods

2.1 Reagents

2.1.1 Buffers

- FSB: 10 mM CH_3COOK, 45 mM $MnCl_2$ (*BDH*), 10 mM $CaCl_2$ (*BDH*), 3 mM $Co(NH_3)_6Cl_3$, 10% [v/v] glycerol (*Fischer Bioreagents*), pH 6.4
- Rich liquid media: 2% [w/v] tryptone (BD), 1.5% [w/v] yeast extract (BD), 0.8% [w/v] NaCl (*VWR International*), 0.2% [w/v] Na_2HPO_4 (*Applichem*), 0.1% [w/v] KH_2PO_4 (*VWR International*) and 0.8% [v/v] glycerol
- LB media: 1% [w/v] tryptone (*BD*), 0.5% [w/v] yeast extract (*BD*), 1% [w/v] NaCl (*VWR International*)
- TAE buffer: 2 M Tris base, 6% acetic acid [v/v], 50 mM EDTA
- 4× P0: 20 mM Tris-acetate pH 7.9, 14 mM 2-mercaptoethanol (*Sigma Aldrich*), 10 mM Mg-acetate (*Acros Organics*), 100 µM $ZnSO_4$ (*VWR International*). 4× P0 buffer was used as a base for many buffers. P buffers denote the addition of potassium acetate. N buffers denote the addition of sodium chloride. For example, P300 is 1× P0 buffer containing 300 mM potassium acetate. N500 is 1× P0 containing 500 mM NaCl.
- SDS PAGE 2× loading buffer: 4% [w/v] SDS, 125 mM Tris HCl pH 6.8, 20% [v/v] glycerol, 0.2% [w/v] bromophenol blue (*Fischer Biotech*) and 4% [v/v] 2-mercaptoethanol
- SDS PAGE running buffer: 25 mM Tris, 190 mM glycine, 0.1% [w/v] SDS, pH 8.3
- Coomassie stain: 60% [v/v] methanol (*VWR International*) and 0.2% [w/v] Brilliant Blue 250 R (*Fischer Scientific*)
- SDS PAGE Destain: 25% [v/v] ethanol (*Sigma Aldrich*), 8% [v/v] acetic acid (*Fischer Bioreagents*)
- 6× DNA loading dye: 10 mM Tris-HCl pH 7.6, 0.03% [w/v] bromophenol blue, 0.03% [w/v] xylene, 60% [v/v] glycerol, 60 mM EDTA
- 10× M9 salts: 0.4 M Na_2HPO_4, 0.2 M KH_2PO_4, 0.1 M NaCl, 0.2 M NH_4Cl
- 100× glucose/ampicillin stock: 20% [w/v] glucose, 10 mg/ml ampicillin sodium salt

- PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4
- 2000× Trace Elements: Per L contained 9 g $Na_2B_4O_7.10H_2O$, 0.44 g $ZnSO_4.7H_2O$, 0.1 g $CuCl_2.2H_2O$, 0.06 g $NaMoO_4.2H_2O$, 0.06 g $VOSO_4.2H_2O$, 0.02 g $CoSO_4.7H_2O$, and 3.6 g $MnCl_2.4H_2O$
- HMNE buffer: 40 mM HEPES pH 7.3, 250 mM NaCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 5% [v/v] glycerol and 10 mM dithiothreitol

2.1.2 Competent Cells

α-select Bronze Efficiency (*Bioline*) and DH5α Subcloning Efficiency (*Invitrogen*) were used for cloning and DNA preparation. One Shot BL21 Star (DE3) (*Invitrogen*) and Rosetta 2 were used for protein expression.

The cloning competent cells have been engineered to produce large amounts of high quality plasmid DNA. The α-select and DH5α cells contain a mutation in *endA1*, which encodes an intracellular endonuclease that can degrade plasmids. This results in higher yields of plasmid preparations. The α-select and DH5α cells also contain a mutation in *recA* to prevent homologous recombination occurring, which may otherwise lead to the disruption of the plasmid of interest.

The expression strains have been engineered to promote high levels of protein expression. BL21 DE3 cells lack *lon* and *ompT*, which encode proteases. This results in higher final yields of purified recombinant protein. DE3 cells also have a chromosomal copy of the T7 RNAP gene, under the control of the *lacUV5* promoter. The operator region of this promoter is bound by the *lac* repressor which prevents expression of the polymerase. Upon the addition of IPTG, the repressor dissociates from the operator and the T7 polymerase is expressed. This enables the specific induction of recombinant protein expression from pET vectors where the gene of interest is positioned downstream of a T7 promoter. Rosetta 2 cells are derivatives of BL21 DE3 cells which also encode tRNAs for 7 rare codons. This facilitates the expression of proteins from species that have a different codon bias to *E. coli*.

2.1.3 Vectors

Expression Three plasmids were used for the expression of protein. Untagged proteins were expressed in pET-21a(+). His-tagged proteins were expressed in pET151D or pET-21a(+). GST-tagged proteins were expressed in a pGEX-2TK vector.

Subcloning and *in vitro* transcription pGEM-T vectors were used for subcloning and as a vector for *in vitro* transcription of riboprobes.

pET vectors contain a T7 promoter, transcription start site, and terminator. This enables expression of genes by T7 RNA polymerase. pET-21a(+) and pET151D vectors also confer resistance to ampicillin, enabling selection of transformants. pET-21a(+) also provides the option of a C-terminal hexa-histidine tag whilst pET151D vectors result in proteins with an N-terminal hexa-histidine tag. In pET vectors, a *lac* operator is located 5′ to the cloned gene. This is bound by the *lac* repressor, blocking transcription. When IPTG is added, it binds to the *lac* repressor, which then dissociates from the operator enabling transcription of the gene. In BL21 DE3 cells, expression of T7 RNAP is also regulated by the *lac* repressor, conferring a double lock on expression of the cloned gene, until expression is induced.

pGEX-2TK vectors are used to express proteins with an N-terminal GST tag. The GST fusion gene is under the control of a *tac* promoter, and is regulated by the *lac* repressor. Addition of IPTG enables expression of the gene. The *tac* promoter is transcribed by the *E. coli* RNAP, which is constitutively expressed. There is therefore only one lock on leaky expression, compared to pET vectors where there are two. The pGEX-2TK vector also confers resistance to ampicillin allowing for selection of transformants. A thrombin-cleavage site between the GST and recombinant protein of interest enables them to be easily separated.

pGEM-T vectors are commonly used for cloning PCR products. It is a linearised vector which contains 3′ T overhangs at both ends. This allows easy ligation from PCR products that have been synthesised by *Taq* DNA polymerase, which leaves A overhangs on the 3′ end of amplified DNAs. pGEM-T vectors confer resistance to ampicillin. They contain T7 and SP6 promoters and are suitable for *in vitro* transcription to produce riboprobes.

2.2 Primers and Oligos

Name	Sequence 5′–3′	Function
KH1 GDDG Fw	GTAGGAGCGGCAATTGGGGATGATGGAGAGAACGTTAAAAC	Removal of RNA-binding activity from NusA KH1 domain. Mutate K50, G51 to D50, D51
KH1 GDDG Rv	GTTTTAACGTTCTCTCCATCATCCCCAATTGCCGCTCCTAC	Removal of RNA-binding activity from NusA KH1 domain. Mutate K50, G51 to D50, D51
KH2 GDDG Fw	GTTAGAAGAGCGGTTTTTGGAGATGATGGCAAGAACTTAGAAAGAGC	Removal of RNA-binding activity from NusA KH2 domain. Mutate E117, K118 to D117, D118
KH2 GDDG Rv	GCTCTTTCTAAGTTCTTGCCATCATCTCCAAAAACCGCTCTTCTAAC	Removal of RNA-binding activity from NusA KH2 domain. Mutate E117, K118 to D117, D118
T7 Promoter Fw	TAATACGACTCACTATAGGG	Upstream primer for splice-by-overlap extension PCR
T7 Terminator Rv	GCTAGTTATTGCTCAGCGG	Downstream primer for splice-by-overlap extension PCR
Mja NusA K64E, K65E Fw	GCAGAAGAGAAATTTGGAGAGGAGGTTGATATTATTGAGTAC	Charge reversal of NusA K65 and K65
Mja NusA K64E, K65E Rv	GTACTCAATAATATCAACCTCCTCTCAAATTTCTCTTCTGC	Charge reversal of NusA K65 and K65
Mja 16S rRNA(A) Leader Fw	TGTTCTATTCCTAAAAACGTTGCATATAAC	Cloning of rRNA leader from *M. jannaschii* genomic DNA (position 159,635–159,460 on negative strand)
Mja 16S rRNA(A) Leader Rv	TAAGTGGGAGGTACGGTCGCAAGAAAGG	Cloning of rRNA leader from *M. jannaschii* genomic DNA (position 159,635–159,460 on negative strand)
Mja 16S rRNA(A) Downstream Fw	TGAGAAAAAAGCGCTGGTTGGTCG	Cloning of rRNA downstream region from *M. jannaschii* genomic DNA (position 157,990–157,910 on negative strand)
Mja 16S rRNA(A) Downstream Rv	ACTTTCCCTATGAGGCA	Cloning of rRNA downstream region from *M. jannaschii* genomic DNA (position 157,990–157,910 on negative strand)
Mja 16S rRNA(A) Leader FUS Rv	TCTCATAAGTGGGAGGTACGGTCGCAAGAAAGG	Cloning of rRNA leader-downstream fusion
Mja 16S rRNA(A) Downstream FUS Fw	CACTTATGAGAAAAAAGCGCTGGTTGCTGC	Cloning of rRNA leader-downstream fusion
Mja NusG Fw	GGGCATATGATTTTTGCAGTTAGAACTATG	Cloning of His-tagged Spt5 mutants from pGEX vector
Mja NusG His Rv	CCCCTCGAGATCTTTATGCTTTGAAAC	Cloning of His-tagged Spt5 mutants from pGEX vector
A2 RNA Cy3	Cy3-CUCUGUUCGAAUUACCAACCUUGCGUG	Fluorescently labelled RNA used in NusA bandshift assays

Table 2.1: Primers and oligos used in this thesis.

2.3 Experimental Methods

2.3.1 Polymerase Chain Reaction

For each 50 μl PCR reaction the following reagents were incubated together on ice in a PCR tube:

- Phusion High-fidelity DNA polymerase (*NEB*): 1 μl
- 5× Phusion HF buffer (*NEB*): 10 μl
- 25 mM dNTP mix (*Sigma Aldrich*): 4 μl
- Forward primer, 10 μM: 2 μl
- Reverse primer, 10 μM: 2 μl
- DNA template, 100 ng/μl: 1 μl
- H_2O: 30 μl

The reaction was conducted in a *Bio-Rad* C1000-Thermo Cycler. The lid was heated at 95°C to prevent condensation of the reaction mixture. The reaction was initially heated to 95°C for 60 seconds. Then a three-step cycle was performed. Within this cycle, the reaction was first heated to 95°C for 30 seconds to melt the DNA strands. Then it was lowered to 5°C below the primer melting temperature for 30 seconds to allow the annealing of the primers to their complementary region on the template DNA. The temperature was raised to 72°C to allow polymerase extension. Generally 30 seconds per kb was given for the extension. This cycle was carried out 35 times in total. The sample was then heated to 72°C for 300 seconds before being held at 12°C. PCR products were purified using a QIAprep PCR purification kit (*Qiagen*) according to the manufacturer's instructions.

2.3.2 Restriction Digests of DNA

Restriction digests were performed using either *Fermentas* FastDigest, or *NEB* enzymes. Approximately 500 ng DNA was digested. 4 µl of 10× FD buffer (*Fermentas*) or 10× CutSmart buffer (*NEB*) was incubated with 1 µl of each restriction endonuclease used in the digestion. The volume was made up to 40 µl with dH_2O. The reaction was incubated at 37°C for 1 hour. 5 μl of the reaction was analysed on an agarose gel. The remaining reaction was purified using a QIAprep reaction cleanup kit (*Qiagen*) according to the manufacturer's instructions.

2.3.3 DNA Ligation into pET Vector

100 ng pET-21a(+) was digested for 1 hour with *EcoRI* (*Fermentas*) and *NdeI* (*Fermentas*) in a 40 µl reaction and treated with 1 µl alkaline phosphatase (*Promega*). The insert was also digested with *EcoRI* and *NdeI*, but not treated with alkaline phosphatase. The restriction enzymes were inactivated by heating the reaction to 65°C for 15 minutes. The digested insert and vector were incubated together in a molar ratio of 6:1 in a total volume of 10 μl, also containing 1 unit DNA ligase (*NEB*) and 1 µl 10× ligase buffer (*NEB*). The reaction was incubated at room temperature for 1 hour and 2.5 µl was transformed into α-select cells. A reaction containing no insert was carried out in parallel as a control.

2.3.4 DNA Agarose Gel Electrophoresis

1% or 0.5% agarose (*Sigma*) gels were made with TAE buffer. The agarose-TAE was melted in a microwave and ethidium bromide (final 0.01% v/v) was added. The agarose was poured into a gel caster, a comb was added, and the gel left at room temperature to set. DNA samples were mixed with 6× DNA loading buffer and loaded into the wells. The gel was run in TAE buffer at 100 V for approximately 30 minutes. A standard (Hyperladder by *Bioline* or Low Range Gene Ruler by *Thermo Scientific*) was included to enable sizing of bands.

2.3.5 Preparation of Competent Cells

An overnight culture of *E. coli* cells was grown in LB, from a single colony. 1 ml of culture was then diluted with 100 ml medium and grown in an incubator at 37°C, shaking at 200 rpm until an OD_{600} of 0.5-0.6 was reached. The cells were then kept on ice for 10 minutes. The cells were centrifuged at 3500 × g for 10 minutes, and the supernatant removed. The pellet was then resuspended in 20 ml FSB buffer, and incubated on ice for 10 minutes. The cells were centrifuged at 3500 × g for 10 minutes and resuspended in 8 ml ice-cold FSB buffer. 280 µl DMSO (*Sigma Aldrich*) was added and the cells incubated for 15 minutes on ice. A further 280 µl DMSO was added and the cells aliquoted into 50 μl fractions in microcentrifuge tubes which had been pre-chilled at -20°C. The cells were snap frozen in liquid nitrogen and stored at -80°C.

2.3.6 Transformation of Competent Cells

Aliquots of competent cells were thawed on ice. 100 ng of plasmid DNA was added and the cells kept on ice for 20 minutes. They were then heat-shocked at 42°C for 45 seconds and put on ice for a further 2 minutes. The cells were then plated onto an LB plate containing ampicillin (*Sigma Aldrich*) at 50 µg/ml and grown overnight at 37°C to enable selection of transformed cells.

2.3.7 Colony PCR

Colony PCR was performed after transformation of ligation reactions to test for the presence of the desired plasmid containing the correct insert. 10 μl of 2× RedTaq ready mix (*Sigma Aldrich*) was incubated with 8 μl H_2O, 1 μl T7 forward primer (final 0.5 μM), and 1 μl reverse primer (final 0.5 μM). A sterile pipette tip was used to pick up a single colony and the tip was placed into the reaction mixture and then plated on an LB-agar plate. The PCR settings were the same as for other PCR reactions with the exception that the initial melting at 95°C lasted for a duration of 10 minutes to enable sufficient disruption of the bacterial cells and liberation of the plasmid template DNA.

2.3.8 Plasmid Purification

A single colony from a transformation of the desired plasmid was added to 5 ml LB media containing 50 µg/ml ampicillin. The culture was grown overnight at 37°C. The plasmid DNA was purified using a QIAprep spin mini-prep kit (*Qiagen*) according to the manufacturer's instructions.

2.3.9 DNA Sequencing

10 μl 100 ng/μl purified plasmid DNA was sent to *Source BioScience* for sequencing, using a T7 forward primer.

2.3.10 Overnight Cultures

A single colony was added to 50 ml liquid media in an Erlenmeyer flask. 500 µl glucose (*Fischer Scientific*) was added at a final concentration of 0.2% [w/v] and ampicillin at a final concentration

of 50 μg/ml. The cells were grown overnight at 37°C, shaking at 200 rpm.

2.3.11 Expression of Recombinant Protein

1 L of rich liquid medium was added to an Erlenmeyer flask. Glucose was added at a final concentration of 0.2% and ampicillin at a final concentration of 50 μg/ml. 15 ml of starter culture, grown the night before, was used to inoculate the media. The culture was grown at 37°C, shaking at 200 rpm until an OD_{600} of 0.6-0.8 was reached. Recombinant protein expression was induced by addition of IPTG (*Glycon Biotechnology*) at a final concentration 1 mM and the cells incubated for a further 5 hours. The cells were centrifuged for 15 minutes at 6000 × g and frozen at -20°C.

2.3.12 Extraction of Protein

Cell pellets were resuspended in N500 buffer. Approximately 15 ml of buffer was used per litre of culture grown. 10 μl of 1000 units/ml DNaseI (*Sigma Aldrich*) and 10μl of 20 mg/ml RNaseI (*Fermentas*) were added. The resuspended pellets were sonicated with 20 seconds on/20 seconds off pulses for 10 minutes, whilst sitting on ice. Cell extracts were then centrifuged at 16,000 × g for 30 minutes at 4°C to separate soluble and insoluble protein.

2.3.13 Heat Inactivation of Supernatants

Supernatants were separated into 1.5 ml aliquots and heated at 65°C for 30 minutes. Samples were then centrifuged at 20,000 × g for 10 minutes, and the insoluble material was discarded.

2.3.14 Ammonium Sulphate Precipitation

Proteins that were to be purified by size exclusion chromatography were first heat-inactivated. The supernatants were placed on ice, and ammonium sulphate was gradually added until it reached saturation. The solution was allowed to sit for 10 minutes and the precipitated protein was removed, with care taken not to remove any undissolved ammonium sulphate. The precipitated protein solution was centrifuged at 16,000 × g for 30 minutes, and the supernatant carefully removed and discarded. The pellet was resuspended in 5 ml buffer, and passed through a 0.22 μm filter.

2.3.15 Protein Purification

N.B. "Wash buffer" refers to the buffer that the protein was extracted or stored in before loading on to the column. "Elution buffer" refers to the buffer that was passed through the column to specifically elute the bound protein of interest.

Non-tagged Proteins

The heat-stable supernatant was ammonium sulphate precipitated, and resuspended in 5 ml buffer. The protein was purified on a HiPrep 16/60 Sephacryl S-100 column (*GE Healthcare*) using a *Bio-Rad* BioLogic DuoFlow FPLC. 5 ml fractions were collected and analysed by SDS PAGE.

His-tagged Proteins

A 1 ml HisTrap FF crude column (*GE Healthcare*) was equilibrated in elution buffer, containing 250 mM imidazole, to remove any previously bound protein. It was then equilibrated in wash buffer containing 20 mM imidazole for five column volumes. The sample was loaded on to the column at a flow rate of 1 ml/min and the flow was collected for analysis by SDS PAGE. After the sample was fully loaded, ten column volumes of wash buffer was passed through the column. The protein was purified in an imidazole gradient, which went from 20 mM to 250 mM imidazole in 10 ml. Then 10 ml 100% elution buffer was passed through the column. The eluent was collected in 1 ml fractions. Wash buffer was passed through the column again for five column volumes to equilibrate it for the next run. After use the columns were washed with water then 20% ethanol and stored at 4°C.

Heparin Affinity purification

Heparin is a highly negatively charged glycosaminoglycan which is able to mimic the structure of DNA and RNA. This makes it a useful molecule for the purification of nucleic acid-binding proteins. A 1 ml HiTrap Heparin HP column (*GE Healthcare*) was equilibrated in elution buffer containing 1000 mM NaCl or NH_4Cl to remove any previously bound protein. It was then equilibrated in wash buffer containing 50 mM NaCl or NH_4Cl for five column volumes. The sample was loaded on to the column at a flow rate of 1 ml/min and the flow was collected for

analysis by SDS PAGE. After the sample was fully loaded, ten column volumes of wash buffer was passed through the column. The protein was eluted in a salt gradient, which went from 50 mM to 1000 mM salt in 10 ml. Then 10 ml 100% elution buffer was passed through the column. 1 ml fractions were collected. Wash buffer was passed through the column again for five column volumes to equilibrate it for the next run. After use the columns were washed with water then 20% ethanol and stored at 4°C.

GST-tagged Proteins

A 1 ml Bioscale mini pro affinity GST cartridge (*Bio-Rad*) was equilibrated in elution buffer, containing 20 mM reduced glutathione, to remove any previously bound protein. It was then equilibrated in wash buffer for five column volumes. The sample was loaded on to the column at a flow rate of 1 ml/min and the flow was collected for analysis by SDS PAGE. After the sample was fully loaded, ten column volumes of wash buffer was passed through the column. Then elution buffer was passed through the column for ten column volumes and the eluent was collected in 1 ml fractions. Wash buffer was passed through the column again for five column volumes to equilibrate it for the next run. After use the columns were washed with water then 20% ethanol and stored at 4°C.

2.3.16 Removal of the GST Tag

Purified GST-tagged proteins were treated overnight with thrombin at 4°C to separate the desired archaeal protein from the GST tag. Samples were then heat treated at 65°C for 30 minutes, resulting in the precipitation of the GST. The heat-stable archaeal protein remained in solution. Samples were centrifuged at 20,000 × g for 10 minutes to separate the insoluble and soluble protein. The supernatant was carefully removed, aliquoted and stored at -20°C.

2.3.17 Concentration of Protein

Purified recombinant proteins were concentrated, if required, in centrifugal filter units (*Millipore*). Samples were placed inside the units and centrifuged at 6000 × g until the desired volume was obtained.

2.3.18 Determination of Protein Concentration

Protein concentrations were determined using a Qubit 2.0 fluorometer (*Invitrogen*) according to the manufacturer's instructions. Briefly, working solution was prepared by preparing a 199:1 ratio of Qubit protein buffer to Qubit protein dye. A standard curve was generated by incubating 10 µl samples of known concentration with 190 µl working solution. 1–10 µl of protein whose concentration was being determined was incubated with working solution at a final volume of 200 µl. The protein concentration was determined by comparison to the standard curve.

2.3.19 SDS PAGE

Polyacrylamide gels were made from a 40% polyacrylamide solution[1] (*Sigma Aldrich*). The recipe for a 16% gel is shown (Table 2.2). The polyacrylamide, Tris, and H_2O were first mixed together. Next, APS (*Acros Organics*) and TEMED (*Sigma Aldrich*) were added to the resolving gel solution which was quickly poured into 0.75 mm thick Mini-PROTEAN glass plates (*Bio-Rad*), leaving about 1 ml at the top. Isopropanol was carefully poured on top to create an even layer and the resolving gel was left to polymerise. The isopropanol was then washed away. APS and TEMED were added to the stacking gel solution, which was quickly poured on top of the resolving gel. A 10 or 15 well comb was inserted and the gel was left to polymerise. Protein samples were mixed with 2× loading buffer in a 1:1 ratio, heated to 95°C for 3 minutes, and loaded into the wells. Gels were run at 200 V for 45 minutes in Tris/Glycine SDS PAGE running buffer. 5 µl of broad range standard (*Bio-Rad*) was run to allow sizing of the bands. Gels were stained with coomassie stain and then destained in SDS PAGE destain buffer to visualise protein bands.

	Stacking	**Resolving**
40% polyacrylamide solution	4 ml	1 ml
1.5 M Tris HCl, pH 8.8 containing 0.4% SDS	2.5 ml	0 ml
0.5 M Tris HCl, pH 6.8 containing 0.4% SDS	0 ml	2.5 ml
10% [w/v] APS	100 μl	100 μl
TEMED	10 μl	10 μl
H_2O	3.5 ml	6.5 ml

Table 2.2: Recipe to make polyacrylamide gels.

[1] 40% polyacrylamide solutions contained a 37.5:1 ratio of acrylamide:bisacrylamide.

2.3.20 Silver Staining

Gels were fixed for 1 hour in fixing buffer (50% [v/v] methanol, 12% [v/v] acetic acid). 100 μL of 37% formaldehyde (*Fischer Bioreagents*) was also added. The gels were then washed three times for 1 minute in 50% [v/v] ethanol and then once in 0.0125% [w/v] di-sodium-thiosulfate (*Sigma Aldrich*) for 1 minute. The gels were washed three times for 20 seconds in dH_2O. They were stained in 0.1% [w/v] silver nitrate (*Sigma Aldrich*) solution, also containing 100 µl of 37% formaldehyde for 20 minutes. The gels were subsequently washed in dH_2O three times in 30 seconds and developed in 6% [w/v] sodium carbonate (*BDH*). When the desired level of development was reached the reaction was stopped with 5% [v/v] nitric acid.

2.3.21 Preparation of Double Spin-labelled Spt5

Double cysteine mutants of Spt5 were expressed with a C-terminal His-tag in a pET-21a(+) vector. The protein was extracted and bound to a NiNTA column. Reducing agent was removed by washing and buffer containing 1 mM iodoacetamide-PROXYL (*Sigma Aldrich*) was passed through the column for 1 hour. The sample was eluted using an imidazole gradient, and further iodoacetamide-PROXYL was added to protein-containing fractions to a final concentration of 3 mM, an approximate 30:1 ratio compared to the protein. The sample was then incubated at 4°C overnight. The sample was diluted 10 times to reduce the imidazole concentration, and was reapplied to a NiNTA column. Unbound spin label was removed by washing. The protein was then eluted in an imidazole gradient.

2.3.22 Ellman's Test

In order to effectively couple spin labels to cysteines it was necessary to ensure there was no reducing agent present. After on-column labelling and purification of double cysteine Spt5 mutants the protein sample was analysed for free thiol groups. 1 µl of 5 mg/ml DTNB reagent (*Sigma Aldrich*) dissolved in methanol was added to 39 µl of the protein preparation. DTNB, which absorbs at 324 nm reacts with thiol groups to produce TNB, which is bright yellow and absorbs at 412 nm. The concentration of accessible thiols was determined using the Beer-Lambert law:

$$A = \epsilon lc \tag{2.1}$$

where ϵ is the extinction coefficient of the protein, l is the path length of light in cm, and c is the molar concentration of the protein.

2.3.23 Dimerisation Between Spt4 and Labelled Spt5

Spt5 and Spt4 were added together in a molar ratio of roughly 1:1.5 and incubated at 65°C for 10 minutes. The sample was then centrifuged for 20 minutes at 20,000 × g. The sample was diluted approximately 10 times in N500 to reduce the imidazole concentration to about 20 mM and purified by Ni-NTA affinity chromatography. Fractions containing Spt4/5 were pooled and concentrated to a final volume of about 200 µl. The sample was diluted two-fold in 50% D8 glycerol (*Sigma Aldrich*) in D_2O and snap frozen in liquid nitrogen.

2.3.24 DEER

All DEER measurements were carried out at 50 K on a *Bruker* ELEXSYS E580 spectrometer at 9 GHz, with an ER-4118-x-MS-3W resonator. Predicted distance distributions were made using rotamer libraries in MMM (version 2013) and the obtained DEER data was analysed using DeerAnalysis (2013 version).

Electronic paramagnetic resonance (EPR) is a spectroscopic technique for analysing paramagnetic centres. Most proteins do not contain a paramagnetic centre, so they must be incorporated through coupling to a spin label. Nitroxide spin labels can be incorporated in a site-specific manner at cysteine residues. Unpaired electrons can exist in two spin states, -1/2 and +1/2, which have different energies when placed within an external magnetic field. An electron in the lower -1/2 spin state can be excited into the higher energy +1/2 spin state if it absorbs microwaves whose energy precisely matches the energy difference between the two spin states. The energy at which an electron is excited can be further influenced by another unpaired electron which is in close proximity due to a phenomenon called dipolar coupling. The strength of the dipolar coupling is proportional to $1/r^3$. Therefore, by determining the strength of the dipolar coupling it is possible to determine the distance between two spin labels on a protein. The experiment that is used to determine the distance distribution between two spin labels is called double elec-

tron electron resonance (DEER). DEER is able to measure distances in the region of 2–8 nm. For a more detailed discussion of EPR techniques, see Chapter 4.

2.3.25 Production of Polyclonal Antibodies

200 μl pure recombinant protein at 1 mg/ml was sent to *Davids Biotechnologie* for immunisation in rabbits. A test serum was taken 35 days after immunisation and tested by ELISA. 40–90 ml final serum was taken after 63 days.

2.3.26 Antibody Purification

200 µl protein G Sepharose resin (*GE Healthcare*) was added to a filter drip column. The protein G resin was equilibrated with 2 ml PBS. 4 ml polyclonal antibody serum (*Davids Biotechnologie*) was added to the column and allowed to clear by gravity flow. The resin was washed in 2 ml PBS and the antibody was eluted by addition of 800 μl 0.1 M glycine, pH 2.8. Then, 80 μl 1 M Tris-HCl, pH 8.5 was added to the purified antibody to neutralise the buffer.

2.3.27 Western Blotting

Recombinant proteins were diluted 100× in PBS, separated by SDS PAGE and transferred to nitrocellulose membranes (*GE Healthcare*) at 100 V for 1 hour on ice in transfer buffer (192 mM glycine, 25 mM Tris base, 20% [v/v] methanol). The membranes were blocked overnight in PBS containing 2% [w/v] BSA (*Sigma Aldrich*) at 4°C. The membranes were then washed three times in PBS containing 0.1% [v/v] Tween 20 (*Sigma Aldrich*). The membranes were subsequently incubated with polyclonal antibody from rabbit serum at a 1:1000 dilution at room temperature for 1 hour. The membranes were then washed again three times for 20 minutes in PBS/0.1% Tween. The membranes were then incubated with anti-rabbit secondary antibody, which was conjugated to DyLight680 (*Thermo Scientific*) for 1 hour at a 1:10,000 dilution. The membranes were washed in PBS/0.1% Tween three times for 20 minutes. They were visualised using a Typhoon FLA 9500 by exciting at 700 nm.

2.3.28 Thermal Denaturation Assay Using Differential Scanning Fluorimetry

20 μl of purified recombinant protein was incubated with 5 µl of SYPRO orange in a 96-well plate. This was placed in a RT PCR machine (*Bio-Rad*). The temperature was increased in 0.5°C increments and the change in fluorescence emission was monitored. The dye was excited at 492 nm and emission was detected at 610 nm.

2.3.29 Limited Proteolysis

25 μg trypsin (*Thermo Scientific*) was resuspended in 250 µl 1 mM HCl, giving a stock concentration of 100 µg/ml. All digests contained a total volume of 15 µl. For single digests recombinant protein (final concentration 40 μM) was incubated with trypsin at the following concentrations in µg/ml: 50, 25, 10, 5, 2.5, 1.125, 0.56, 0.28, 0.14. Digests were left at 37°C for 1 hour. 15 µl 2× SDS loading buffer was added to stop the reaction and 10 µl was analysed by SDS PAGE.

2.3.30 Analytical Size Exclusion Chromatography

Superose 6 and Superose 12 size exclusion chromatography were carried out to analyse interactions between purified recombinant proteins. The column (Superose 6 10/300 GL or Superose 12 10/300 GL) was first washed in two column volumes of filtered dH_2O, followed by two column volumes of filtered buffer. 100 µl of size standard (*Bio-Rad*) was run at 0.5 ml/min to allow sizing of protein complexes (Figure 2.1). Proteins were run individually on the column and sized by comparison to the standard. Proteins were then incubated together at 65°C for 30 minutes, centrifuged at 20,000 × g for two minutes and separated on the Superose column. Proteins/complexes were sized and their identity confirmed by SDS PAGE.

Superose columns are packed with crosslinked agarose-based sugars, which form a mesh containing pores of different sizes. Larger proteins are unable to enter the pores and travel a direct path through the column, and elute first. Smaller proteins are able to enter the pores and travel a longer path through the column, therefore eluting later.

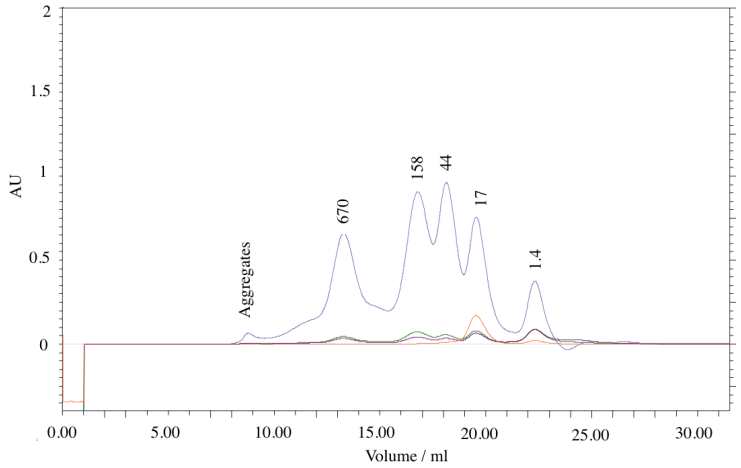


Figure 2.1: Example of a Superose 6 standard run. 100 μl standard was loaded into the the column, and separated at 0.5 ml/min. The larger proteins eluted first as they took a shorter path through the column. The sizes of the standard proteins are 670, 158, 44, 17, and 1.4 kDa. Aggregates result in a small peak at 8.5 ml, which represents the void volume of the column.

2.3.31 Sizing of Putative Antitermination Factors

1.5 g mid-log phase *M. jannaschii* cell mass[2] was ground using a mortar and pestle on ice for 5 minutes with 3 g Al_2O_3 powder (*Sigma Aldrich*). 3 ml lysis buffer (20 mM Tris pH 7.5, 50 mM NH_4Cl, 10 mM $MgCl_2$) was added and the cells were ground for a further 10 minutes on ice. The cell extracts were evenly aliquoted into microcentrifuge tubes. Extracts were then treated with either 10 μl DNaseI, 10 μl RNaseI, 10 μl DNaseI and 10 μl RNaseI, or untreated. The extracts were heated at 37°C for 2 hours. The extracts were then centrifuged twice at 20,000 $\times$ g for 20 minutes at 4°C. 100 µl of the resulting supernatants were separated by Superose 6 size exclusion chromatography and 500 µl fractions were collected. 400 µl of each fraction was added to 100 µl of 6$\times$ SDS PAGE loading buffer. 20 µl samples were separated by SDS PAGE, and Western blot analysis was conducted against the putative antitermination factors. A size standard was also run to enable rough determination of the molecular weight of protein complexes.

[2] *M. jannaschii* cell mass was provided by Prof. Michael Thomm.

2.3.32 Ribosome Preparation from *M. jannaschii* Cell Mass

M. jannaschii Cell Extraction

Mid-log phase *M. jannaschii* cell mass was stored at -80°C. *M. jannaschii* cells and Al_2O_3 powder (*Sigma Aldrich*) were ground using a mortar and pestle on ice for 5 minutes in a mass ratio of 1:2. 10 µl 1000 units/ml DNaseI was added and the cells were ground for a further 5 minutes. Two volumes of lysis buffer (20 mM Tris pH 7.5, 50 mM NH_4Cl, 10 mM $MgCl_2$) was added and the cells were ground for a further 10 minutes on ice.

S30 Fraction

The suspension from the cell grinding was centrifuged at 30,000 × g in a *Beckman* optima L-100 XP Ultracentrifuge in a Type 70 Ti rotor for 30 minutes. The supernatant was carefully removed and was centrifuged again under the same conditions. This removed the cell wall and other insoluble materials.

S100 Fraction

The S30 fraction was processed further to give an S100 fraction. 7 ml S30 supernatant was centrifuged at 100,000 × g for 4 hours through a 9 ml 18% sucrose cushion in a Type 70 Ti rotor, using a *Beckman* optima L-100 XP Ultracentrifuge. The supernatant was carefully removed and the ribosome-containing pellet was resuspended in 2 ml wash buffer (20 mM Tris pH 7.5, 50 mM NH_4Cl, 10 mM $MgCl_2$, 7 mM 2-mercaptoethanol).

Analysis of Ribosome Preparation

Ribosomes were analysed on a sucrose gradient. An 11.5 ml sucrose (*Sigma Aldrich*) gradient was prepared from 10 to 30% sucrose. 3.5 ml 30% sucrose was pipetted into the bottom of the centrifuge tube. 3.5 ml 20% sucrose was pipetted on top with care taken not to disrupt the 30% layer underneath. 3.5 ml 10% sucrose was pipetted on top with care taken again not to disrupt the layers below. The centrifuge tube was left vertically at room temperature for 18 hours to allow a linear gradient to form. 200 µl crude ribosomes was loaded onto the gradient and centrifuged at 62,000 × g for 17 hours in a SW41 Ti rotor. The gradient was then carefully separated in 250 µl fractions by pipetting and 9 µl was analysed on a 1% agarose gel.

2.3.33 Bioanalyzer to Analyse RNA

Ribosomal RNA was extracted from purified ribosomes using an RNeasy kit (*Qiagen*) according to the manufacturer's instructions and eluted in 30 μl RNase-free water. An *Agilent* RNA 6000 Nano Kit was used to analyse 1 μl samples of RNA on a RNA Nano Chip according to the manufacturer's instructions. The RNA Nano Chips were run on an *Agilent* 2100 Bioanalyzer instrument.

2.3.34 Labelling of Ribosomes with Maleimide-Coupled NT650 Dye

The ribosome preparation was labelled with NT650 using an L006 Monolith kit (*NanoTemper*) according to the manufacturer's instructions. Unbound dye was separated from the labelled ribosomes on a gel filtration column, and the labelled ribosomes were eluted in fractions of 100 μl.

2.3.35 Protein Preparation for MST

Spt4/5 variants were expressed as GST-fusions in Rosetta 2 cells. The protein was extracted in MST protein buffer (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 300 mM NH_4Cl, 3.5 mM 2-mercaptoethanol). The Spt4/5 variants were first purified on a GST affinity column and eluted in MST protein buffer containing 20 mM reduced glutathione. The purified GST fusion was treated with thrombin overnight and heated at 65°C for 30 minutes to precipitate the GST. The sample was then centrifuged at 20,000 × g for 5 minutes, the supernatant carefully removed, and diluted 6× in MST protein buffer containing no NH_4Cl. The diluted protein was purified on a heparin affinity column. The protein was loaded on to the column in MST protein buffer containing 50 mM NH_4Cl and eluted in a gradient with MST protein buffer containing 1 M NH_4Cl. Fractions containing Spt4/5 variants were buffer-exchanged to MST protein buffer containing 300 mM NH_4Cl and concentrated in *Millipore* centrifugal units at 6,000 × g.

2.3.36 Generation of NusA KH Domain GXXG–to–GDDG Mutants

The GXXG motif of KH domains, where X is any amino acid, is necessary for RNA binding and its mutation to GDDG abolishes RNA binding without affecting protein integrity (Hollingworth et al., 2012). In order to mutate the NusA KH1 and KH2 domain GXXG loops, a PCR-based

protocol called splice by overlap extension (SOE) was utilised (Heckman and Pease, 2007). For each mutation three separate PCR reactions were performed. pET-21a(+) containing a wild-type *M. jannaschii* NusA insert was used as template DNA. pET-21a(+) vectors contain a T7 promoter and a T7 terminator. The first step in the SOE process required two PCR reactions. The first reaction utilised a T7 forward primer and a reverse primer within the gene that contained the desired mutation. The second PCR reaction used a forward primer containing the desired mutation and a T7 terminator reverse primer. The two reactions yielded overlapping fragments of the final desired PCR product, with the 3′ end of the first PCR reaction and the 5′ end of the second PCR reaction being complementary to each other. The overlapping regions also contained the desired mutation as the mismatched primers were incorporated into the PCR product. The two PCR products were then separated on an agarose gel and purified using a gel extraction kit (*Qiagen*) according to the manufacturer's instructions. The two purified PCR products were then incubated together as template DNA in the third and final PCR reaction, utilising a T7 forward primer and a T7 terminator reverse primer. The template DNAs annealed at the region of mutation and were extended by the DNA polymerase to form one product, containing the desired mutation (Figure 2.2).

An empty pET-21a(+) vector was then digested with the restriction endonucleases *EcoRI* and *NdeI* and treated with alkaline phosphatase to prevent self-annealing. The product of the third PCR reaction was also digested with *EcoRI* and *NdeI*. The digested insert and vector, with complementary sticky ends, were then incubated together in a ligation reaction in a 6:1 ratio, resulting in the insertion of the mutated NusA gene into the pET-21a(+) vector (Figure 2.3).

2.3.37 EMSAs with Fluorescent A2 RNA

A2 RNA with a 5′ Cy3 label at a final concentration of 50 nM was incubated with different concentrations of protein (1, 2.5, 5, 7.5, 10, 15, and 20 μM) in HMNE buffer for 5 minutes at 65°C in a volume of 20 μl. 10 μl 15% ficoll was added and 10 μl of each sample was immediately loaded onto a 12% native polyacrylamide gel, which was run at 100 V for approximately 2 hours. The gel was visualised in a Typhoon FLA 9500 by exciting with a laser at 550 nm.

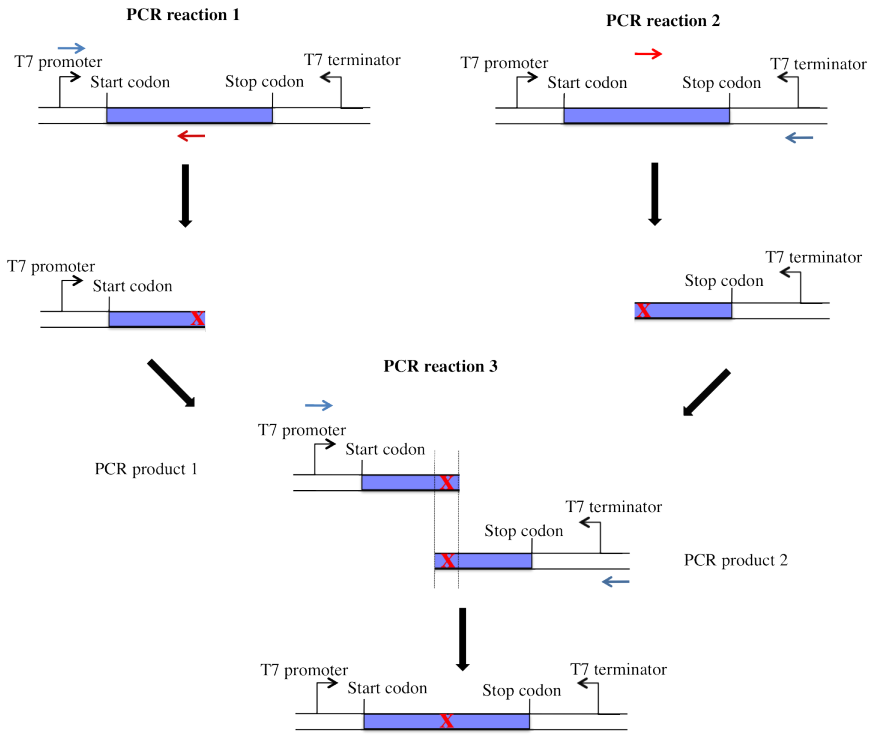


Figure 2.2: Splice by overlap extension for introducing mutations into the *M. jannaschii* NusA gene. Initially, two PCR reactions are carried out, resulting in the production of two overlapping products, each containing the desired mutation and being fragments of the final desired product. The products of the first two PCR reactions are purified and incubated together as template DNA in the final PCR reaction which results in one PCR product.

2.3.38 *In vitro* Transcription of Riboprobes

A pGEM-T vector containing the riboprobe insert was linearised by *PstI* digestion. *In vitro* transcription of riboprobes from the linearised pGEM-T vector was performed using a T7 MAX-Iscript kit (*Ambion*). The following reagents were incubated together at 37°C: 2 µl 10× T7 buffer, 6 µl H_2O, approximately 2 µg template DNA, 1 µl 10 mM ATP (final 500 µM), 1 µl 10 mM GTP (final 500 µM), 1 µl 10 mM CTP (final 500 µM), 2.5 µl $^{32}P\alpha$-UTP (25 µCi), and 2 µl T7 RNAP mix. The reaction solution was incubated for 30 minutes at 37°C and then cold UTP was added to a final concentration of 100 µM. The reaction solution was then incubated for 2 hours at 37°C. 2 µl 1000 units/ml DNaseI was added to digest the template DNA for 15 minutes. 20 µl RNase-free water was subsequently added and the reaction solution heated to

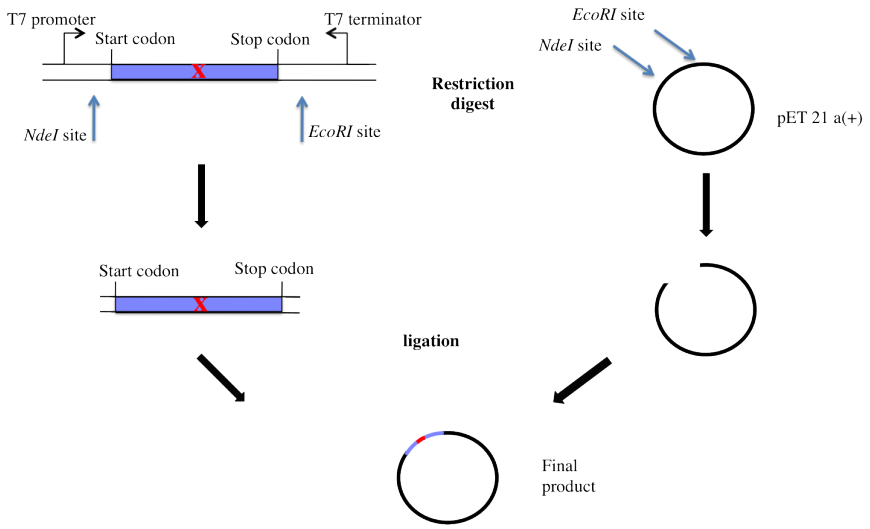


Figure 2.3: Splice by overlap extension for introducing mutations into the *M. jannaschii* NusA gene (continued). The SOE PCR product and an empty pET-21a(+) vector are both digested with *EcoRI* and *NdeI*, yielding two products with complementary sticky ends. The insert and vector are ligated together to yield the final desired construct.

75°C for 2 minutes to heat inactivate the T7 RNAP. The RNA was then purified using a G25 column (*GE Healthcare*) according to the manufacturer's guidelines.

2.3.39 EMSAs with Riboprobes

1:50 dilutions of riboprobes in HMNE were prepared. 10 μl of the diluted riboprobe was incubated with 10 μl protein at varying final concentrations (25 nM, 125 nM, 250 nM, 500 nM, 1μM, 2μM, and 10μM) at 65°C for 5 minutes. The samples were then placed on ice and 10 μl 15% ficoll was added. 10 μl was loaded onto an 8% continuous native gel made from 0.5 M Tris-HCl, pH 6.5. The gel was run at 100 V for approximately 1 hour. The gel was transferred to Whatman paper and dried. It was exposed to an imaging plate for approximately 12 hours and was visualised using a Typhoon FLA 9500 phosphoimager.

2.3.40 Preparation of ^{15}N Minimal Media

860 ml H_2O was autoclaved in an Erlenmeyer flask. To this the following was added: 100 ml sterile 10× M9 salts (minus NH_4Cl), 20 ml of 20% [w/v] sterile glucose, 5 ml of 20% [w/v] $^{15}NH_4Cl$, 100 µl of sterile 1 M $CaCl_2$ (*BDH*), 1 ml of sterile 1 M $MgSO_4$ (*Fischer Scientific*) and 10 ml of 100× MEM salts (*Sigma Aldrich*).

2.3.41 ^{15}N Labelling of Protein

For each litre of expression culture a 25 ml overnight starter culture was grown in rich media. The starter culture was centrifuged at 6000 × g for 15 minutes. The supernatant was removed and the pellet resuspended in 10 ml of sterile 1× M9 salts. The resuspended cells were centrifuged again at 6000 × g for 15 minutes, the supernatant was removed and the cells resuspended in 2 ml of sterile 1× M9 salts and added to 1 L of ^{15}N minimal media. Cells were grown at 37°C shaking at 200 rpm until an OD_{600} 0.6 was reached. IPTG was added at a final concentration of 1 mM. Cells were grown for a further 5 hours at 37°C. The cells were centrifuged for 15 minutes at 6000 × g and the pellets were stored at -20°C.

2.3.42 Preparation of ^{15}N, ^{13}C, ^{2}H Minimal Media

For 1 L media the following was added:

- $^{15}NH_4Cl$: 3 g
- ^{13}C glucose (*Cambridge Isotopes*): 5 g
- Na_2HPO_4: 7 g
- KH_2PO_4: 6.6 g
- $MgSO_4$: 0.5 g
- Na_2SO_4: 0.7 g
- Sodium succinate: 5.4 g
- 10× MEM vitamins: 10 ml
- 2000× trace metals: 500 μl
- Ampicillin: 0.1 g
- 1 M $CaCl_2$: 300 μl
- D_2O (*Sigma*): Up to 1 L

The media was subsequently filter sterilised and stored at 4°C until required.

2.3.43 ^{15}N, ^{13}C, ^{2}H Labelling of Protein

The morning before the labelled protein was prepared, a single transformed colony was isolated in 100 ml LB media and grown at 37°C for approximately 10 hours. The culture was centrifuged

at 6000 × g and resuspended in 20 ml of ^{15}N, ^{13}C, ^{2}H minimal media per litre to be expressed. The resuspended cells were grown at 37°C overnight. 20 ml overnight culture was then added to each litre of media, giving an OD_{600} of approximately 0.1. The cells were grown at 37°C to an OD_{600} of 0.6 and protein expression was induced by the addition of IPTG at a final concentration of 1 mM. After 5 hours of protein expression the cells were centrifuged at 6000 × g for 15 minutes, and the pellets were stored at -20°C.

2.3.44 Scaffold-Independent Analysis (SIA) - Sample Preparation

NusA samples for SIA analysis were expressed in ^{15}N minimal media and extracted in N50 buffer. The soluble fraction was heat inactivated at 75°C for 30 minutes and centrifuged at 20,000 × g for 3 minutes. The heat stable NusA in the supernatant was purified by heparin affinity chromatography. The sample was loaded in wash buffer consisting of 50 mM NaCl and 10 mM Na_2HPO_4, pH 6.5. It was eluted in a gradient consisting of wash buffer and elution buffer (1 M NaCl and 10 mM Na_2HPO_4, pH 6.5). The eluted protein was concentrated and the buffer exchanged to 50 mM NaCl, 10 mM Na_2HPO_4, pH 6.5 (wash buffer) in a *Vivacon* concentrator.

SIA is used to determine the sequence specificity of RNA-binding KH domains (Beuth et al., 2007). KH domains bind regions of RNA which are 5 nucleotides in length, of which the last four are bound in a sequence-specific manner (Lewis et al., 2000). SIA works on the principle that the sequence identity of each nucleotide position can be identified independently of the other nucleotides within the binding motif. This is achieved by preparing ^{15}N-labelled protein, which is subsequently analysed by HSQC spectroscopy in the presence of pools of RNA pentamers. The identity of each position in the pentamer is random (any of A, C, G, or U) except for the position that is being analysed whose identity is known and constant within a given RNA pool. For example, to analyse the sequence preference at position 1, four spectra would be recorded. In the first spectrum, position 1 would always be A. In the second spectrum, position 1 would always be C. In the third spectrum, position 1 would always be G. In the fourth spectrum, position 1 would always be a U. Since the other positions of the RNA pentamers are completely random, this enables the relative preference of the KH domains at each position to be determined independently of the other positions. The nucleotide preference is obtained by determining which nucleotide results in the greatest change in chemical shifts of selected peaks upon the addition

of RNA pentamers compared to the spectrum of free protein. For a more detailed description of SIA, see Chapter 5.

Chapter 3

Characterising Complexes Containing Spt4/5, S10, and NusA

3.1 Introduction

3.1.1 Transcription-Translation Coupling in *E. coli*

Transcription is the most important stage in gene expression and its three main products are all required for translation. The ribosome is composed of rRNA and ribosomal proteins. tRNAs bring amino acids to the 3′ end of the nascent chain through codon-anticodon base-pairing. The mRNA product of transcription is the template for translation (Figure 3.1).

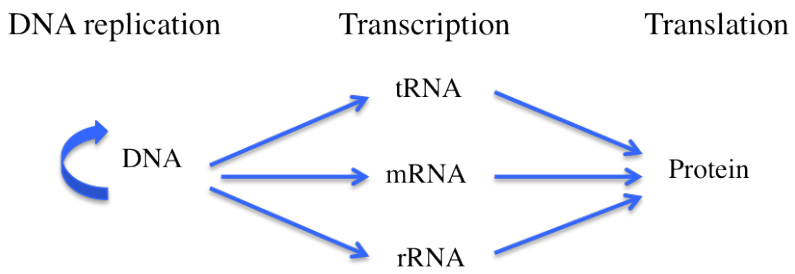


Figure 3.1: The three main products of transcription, tRNA, mRNA and rRNA are all required for translation.

Translation is the process whereby individual amino acids are sequentially incorporated into a polypeptide chain from the N to C terminus by the ribosome. Ribosomes are large macromolecular machines that contain an RNA core, which carries out the catalysis, and ribosomal proteins that stabilise and enhance the catalytic properties of the RNA. In archaea, ribosomes are composed of a 50S and a 30S subunit which associate to form a functional 70S complex. The small subunit contains a 16S RNA and the large subunit contains a 5S RNA and a 23S RNA. Whilst high-resolution structures are available for complete bacterial (Schuwirth et al., 2005) and eukaryotic (Ben-Shem et al., 2011) ribosomes, no such structure currently exists for the archaeal ribosome. However, models are available based on cryoEM data (Armache et al., 2013). It has been known for many years that transcription and translation are co-regulated in bacteria through mechanisms such as attenuation (see Chapter 1). More recently it has become apparent that transcription and translation are directly coupled. NusG is able to bind to RNAP via its NGN domain (Hirtreiter et al., 2010a) and to the ribosome via its KOW domain (Burmann et al., 2010b). *In vivo* data indicates that the transcription and translation rates are correlated,

and that the ratio of amino acids incorporated into the nascent chain to nucleotides added to the RNA is always 3:1. This provides further evidence that transcription and translation are directly coupled (Proshkin et al., 2010).

The ribosome binding face of S10 is still accessible when in complex with NusG. NusG may therefore act as a bridge between transcription and translation complexes (Figure 3.2), bound to the RNAP coiled-coil clamp helices via the NGN domain and to S10 via the KOW domain. RfaH is a NusG paralogue (Bailey et al., 1997) with an N-terminal NGN domain that is structurally similar to that in NusG (Belogurov et al., 2007). Like NusG, it also increases the processivity of RNAP (Artsimovitch and Landick, 2002) by enclosing the DNA within the RNAP active site (Sevostyanova et al., 2011). RfaH has recently been shown to undergo a complete refolding of its C-terminal domain upon binding its target DNA sequence, enabling it to also couple transcription and translation by linking RNAP and the ribosome (Burmann et al., 2012).

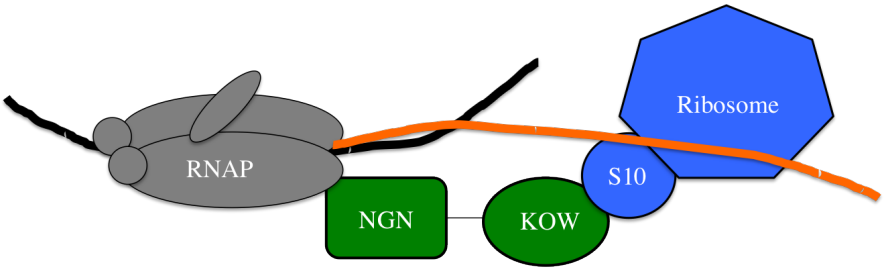


Figure 3.2: NusG is hypothesised to act as a direct protein link between transcription and translation. In addition to being connected by RNA (orange), RNAP and the ribosome are bridged by NusG. NusG binds to RNAP via the NGN domain and to the S10 protein of the small ribosome subunit via the KOW domain.

3.1.2 Chapter Aims

The aim of this chapter is to determine whether the putative antitermination factors Spt4/5 and NusA associate with RNAP and the ribosome in cell extract. In order to test this, preparations of *M. jannaschii* cell lysate were made and separated by size exclusion chromatography. If the factors elute in the same fraction peaks, this will suggest association. Western blot analysis was carried out against the fractions using polyclonal antibodies generated against recombinant

factors. Comparison of elution volumes with a size standard enabled sizing of protein complexes.

Whilst the interaction between archaeal Spt4/5 and RNAP is well documented (Hirtreiter et al., 2010a) the interaction with ribosomes has yet to be characterised. This work in this chapter therefore aims to observe this interaction. This was achieved by performing microscale thermophoresis on ribosomes prepared from *M. jannaschii* cell extract and recombinant Spt4/5. The final aim of this chapter is to ascertain whether stable interactions can be observed between Spt4/5 and recombinant ribosomal protein S10. This was examined by Superose 12 size exclusion chromatography.

3.2 Characterisation of Complexes Containing Spt4/5, NusA, and S10 in Cell Extract

Whilst the interaction between recombinant archaeal Spt4/5 and RNAP is well documented *in vitro* (Hirtreiter et al., 2010a) the interaction has yet to be observed in the context of endogenous proteins from the *M. jannaschii* cell. Furthermore whilst interactions between NusG and recombinant S10 have been detected, interactions between Spt4/5 and translation complexes containing complete ribosomes have not. To determine whether these interactions occur *in vivo*, preparations of mid-log *M. jannaschii* cell extract were made and separated by size exclusion chromatography. Western blot analysis was then carried out against the fractions using polyclonal antibodies generated against recombinant factors.

First, a size standard was run on a Superose 6 column to enable sizing of any complexes (Figure 3.3). Plotting the log of the molecular weight against the elution volume enabled the generation of a standard curve (Figure 3.3B) with the equation:

$$y = -3.26x + 24.4 \tag{3.1}$$

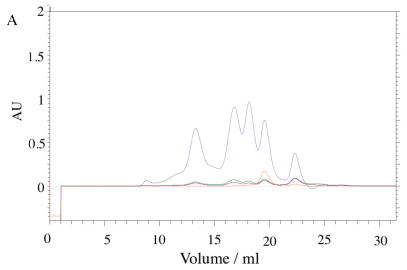
where y is the elution volume and x is log MW. This can be rearranged to:

$$x = (24.4 - y)/3.26 \tag{3.2}$$

so that for any given elution volume, the molecular weight can be calculated.

There is a small peak at 8.7 ml (Figure 3.3A). This represents the exclusion limit of the column. The accuracy of size-determination decreases as the elution volume approaches the limit of the column.

A



B

Size standard (kDa)	Log MW	Elution volume (ml)
670	5.83	13.21
158	5.20	16.71
44	4.64	18.16
17	4.23	19.57
1.4	3.13	22.32

C

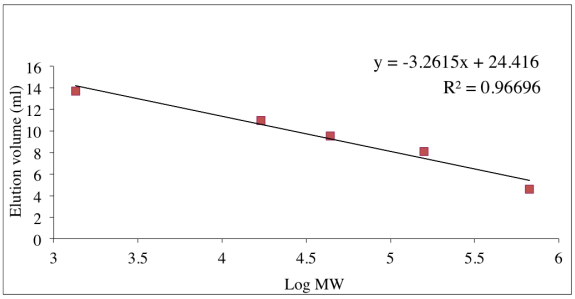


Figure 3.3: Superose 6 size standards. A) Chromatogram of the standard run. B) Elution volumes of the five proteins in the size standard. C) Plotting the elution volume of the proteins in the size standard against log MW enables the generation of a standard curve. Using the standard curve the molecular weight of any complex can be determined, given its elution volume.

Due to the nucleic acid binding properties of the putative antitermination complexes, *M. jannaschii* cell extract was also treated with DNaseI or RNaseI for 1 hour at 37°C. The nuclease-treated cell extracts were analysed on a 1% agarose gel (Figure 3.4A). After DNase treatment, the largest nucleic acids have been destroyed and there is a very strong band at 600 bp. RNase treatment results in the disappearance of the intense band at 600 bp. Treating the extract with DNase and RNase removes nearly all of the high molecular weight nucleic acids and the intensity on the gel is significantly less than for the DNase-only and RNase-only treated extracts (Figure 3.4A). The nuclease-treated extracts were also separated by Superose 6 size exclusion chromatography. Western blots were carried out against eluted fractions to determine whether the complexes are nuclease-sensitive. After DNase treatment there is a slight difference in the

Superose 6 chromatogram (Figure 3.4B&C). The peak at 21.5 ml has a greater intensity for both protein (280 nm) and nucleic acid (260 nm). After RNase treatment (Figure 3.4D), the 260 nm and 280 nm absorbance at 17 ml is greatly reduced compared to untreated extract.

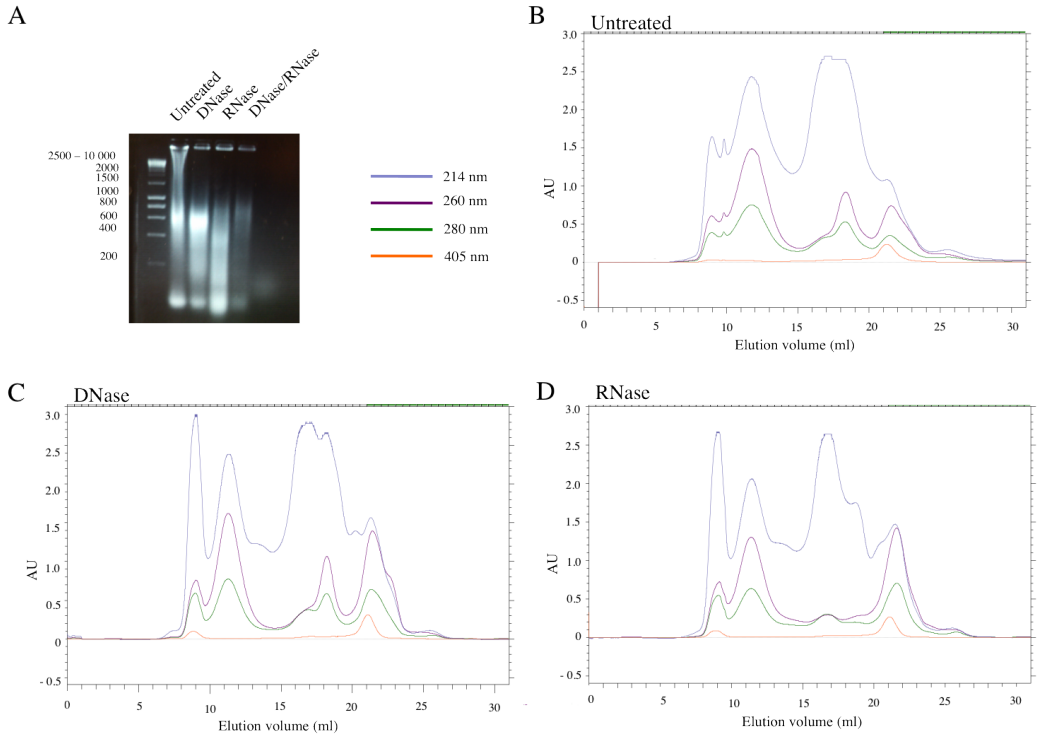


Figure 3.4: Chromatograms of Superose 6 separation of *M. jannaschii* cell extract. A) 10 μl nuclease-treated *M. jannaschii* cell extract was resolved on a 1% agarose gel. B) Untreated extract. C) DNase-treated extract. D) RNase-treated extract. RNase treatment results in the disappearance of the 260 and 280 nm peaks at 17 ml. Light blue: 214 nm, purple: 260 nm, green: 280 nm, orange: 405 nm.

3.2.1 Analysis of Rpo4/7 Elution Profile from *M. jannaschii* Extract

After fractionation of cell extract, Western blot analysis was performed using antibodies against the RNAP subunits Rpo4/7 which form the stalk (Figure 3.5A). The stalk was chosen because, in addition to being a stably incorporated sub-complex of the RNAP (Grohmann et al., 2009), it binds to the emerging RNA transcript (Hirtreiter et al., 2010b; Meka et al., 2005). RNAP elutes in three peaks when the cells are untreated. A high molecular weight complex at 10 ml

represents a mass in the MDa range. This likely corresponds to transcription complexes that are coupled to translation complexes. A peak at 13-14 ml corresponds to RNAP that is not bound to ribosomes with a molecular weight of approximately 470 kDa. *M. jannaschii* RNAP has a molecular mass of approximately 370 kDa. However, given that it is bound to DNA and is likely to be associated with additional transcription factors it is not unexpected to find it in complexes of higher molecular weight. The third peak at 18-20 ml corresponds to a molecular weight of just under 20 kDa, which is less than the actual Rpo4/7 molecular weight of 33.5 kDa. The Rpo4/7 that elutes here represents free subunits that have either not been incorporated into RNAP or have dissociated from the RNAP during extraction of the protein from the cell. Since previous data suggests that Rpo4/7 is stably incorporated into RNAP, it is more likely to be excess Rpo4/7 that has yet to be incorporated into an RNAP.

Upon treatment of the cell extract with DNaseI the peak corresponding to ribosome-associated RNAP in the MDa range disappears (Figure 3.5B). This could be because RNAP was bound to large regions of genomic DNA. Alternatively, DNase treatment may have broken up large RNAP-protein complexes that are dependent on DNA for their integrity. The other two Rpo4/7 elution peaks remain; there is still a peak at 13-15 ml which represents transcription complexes and a peak at 18-20 ml representing the free subunits.

Treatment of the cell extract with RNaseI abolishes the highest molecular weight complexes of RNAP at 10 ml, although some ribosome-bound RNAP is still present at 11-12 ml (Figure 3.5C). RNase treatment does not affect the RNAP fraction, with Rpo4/7 still eluting at 13-15 ml. However, the peak that represents free Rpo4/7 has upshifted in its profile, eluting at 16-19 ml, compared to 17-20 ml for non-treated extract. This could be because large RNAs are degraded into shorter RNAs that can now be bound by Rpo4/7, resulting in an increase in RNAP mass. Each nucleotide in an RNA molecule is approximately 330 Da on average, so every three nucleotides would add 1 kDa to the mass of a complex.

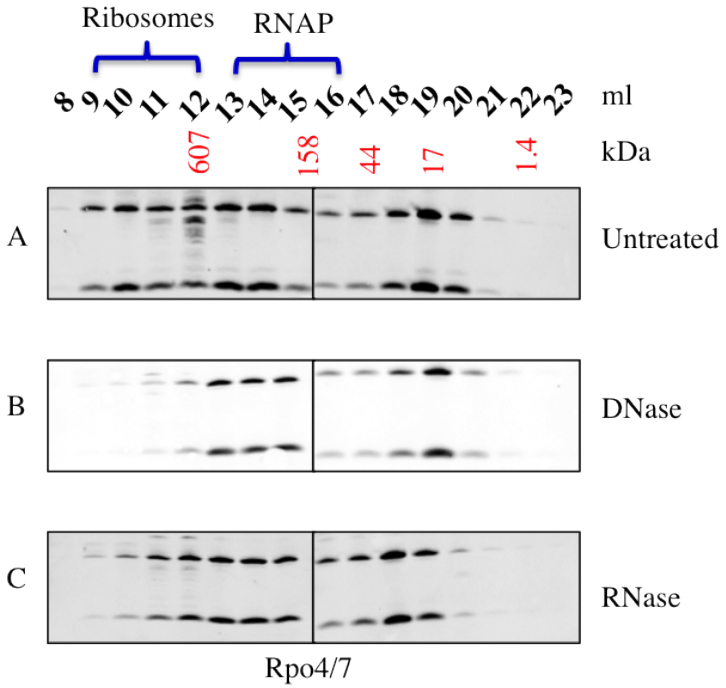


Figure 3.5: Western blot analysis of Superose 6 separated Rpo4/7 from cell extract. Comparison with a size standard (red) enabled sizing of complexes. A) In untreated extract Rpo4/7 elutes in three peaks corresponding to ribosome-bound, transcription complexes and free subunits. B) Upon DNase treatment of the cell extract Rpo4/7 is no longer found associated to ribosomes. C) After RNase treatment, Rpo4/7 does not elute in the highest molecular weight fractions. The elution volume for free Rpo4/7 is upshifted by 1 ml.

3.2.2 Analysis of S10 Elution Profile from *M. jannaschii* Extract

Western blots were also carried out on the same separated *M. jannaschii* cell extract using antibodies against the ribosomal protein S10 (Figure 3.6A). There is a peak from 10-13 ml representing a complex in the MDa range. This is the ribosomal fraction. There is no peak at 13–14 ml, which corresponds to the mass of RNAP. Therefore S10 does not associate with RNAP. There is also no peak at 19-20 ml corresponding to the mass of free S10, 12.24 kDa. Therefore there is no free S10, and S10 is present only in ribosome-bound form. *E. coli* S10 has been reported to have low solubility by itself *in vitro* when not ribosome-bound (Luo et al., 2008). It is therefore possible that any free *M. jannaschii* S10 precipitates out of solution during

the cell extract preparation. Consequently no free S10 would be loaded on to the Superose 6 column, and therefore no free S10 would be detected.

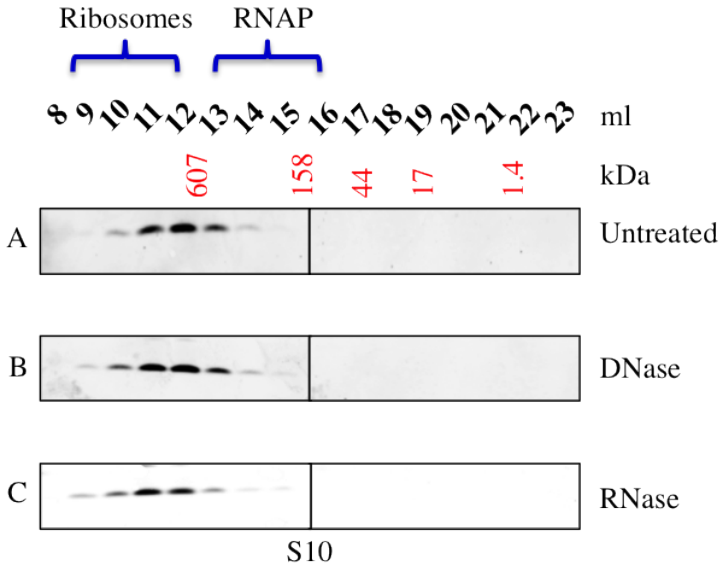


Figure 3.6: Western blot analysis of Superose 6 separated S10 from cell extract. Comparison with a size standard (red) enabled sizing of complexes. A) In untreated extract S10 elutes in one peak only, corresponding to it being incorporated into the ribosome. B) DNase treatment of the cell extract has no effect on the elution profile of S10. C) RNase treatment of the cell extract has very little effect on the elution profile of S10. No free S10 is observed.

Treatment with DNaseI does not disrupt the complex (Figure 3.6B). This is not unexpected since ribosomes are composed mainly of RNA. In addition there is no reported evidence of S10 binding to DNA. However, RNase treatment does not greatly alter the elution profile either (Figure 3.6C), which demonstrates that the ribosomes are still completely intact and are therefore highly stable and resistant to RNase treatment. This may be due to the ribosomal proteins protecting the rRNA from degradation. After both DNase treatment and RNase treatment, the S10 at 10 ml becomes more prominent and it also becomes detectable at 9 ml.

3.2.3 Analysis of Spt4/5 Elution Profile from *M. jannaschii* Extract

Spt4/5 elutes from the size exclusion column in three peaks (Figure 3.7A). Only Spt5 is visible on the Western blots suggesting that very few of the antibodies in the polyclonal sample bind Spt4, and that Spt4 is not very immunogenic. At 10-11 ml there is a peak corresponding to a very high molecular weight complex of about 5.5 MDa, based on the size standard. This overlaps with the elution of S10. This fraction is therefore likely to be ribosome-associated Spt4/5. Ribosomes have a mass of 2.5 MDa in bacteria and 4.6 MDa in eukaryotes. Archaeal ribosomes are not larger than their eukaryotic counterparts. However 10 ml is approaching the void volume of the column (8.7 ml)(Figure 3.3A) explaining the poor accuracy of size prediction. The peak from 14-15 ml corresponds to a molecular weight of approximately 470 kDa. This fraction is most likely transcription complexes as it overlaps with the elution of RNAP. The third Spt4/5 peak is at 18-20 ml. Using the size standard, an elution volume of 19 ml corresponds to a molecular weight of 20 kDa. Spt4/5 has a mass of 24 kDa. Therefore this low molecular weight peak is free Spt4/5 that is not in complex with any other proteins. This data suggests that in the context of *M. jannaschii* cell extract, Spt4/5 is found bound to ribosomes, bound to RNAP, and in its free state. When also considering that RNAP is found in the ribosome fractions, this raises the possibility that Spt4/5 couples transcription and translation in archaea.

Two species of Spt5 are detected on the Western blot of untreated extract which have slightly different electrophoretic mobilities. The RNAP-associated Spt5 at 14-15 ml has a slightly higher mass than ribosome-bound and free Spt5 (Figure 3.7A). This suggests that transcription complex associated Spt4/5 may be post-translationally modified. In eukaryotes Spt4/5 is phosphorylated. Post-translational modifications are not as common in archaea as in eukaryotes. However, they have been reported in *M. jannaschii* (Giometti et al., 2002), including acetylation and methylation (Forbes et al., 2004). Isolating Spt4/5 from cell extract by immunoprecipitation and analysing it mass spectroscopy would enable identification of any post-translationally modified Spt5.

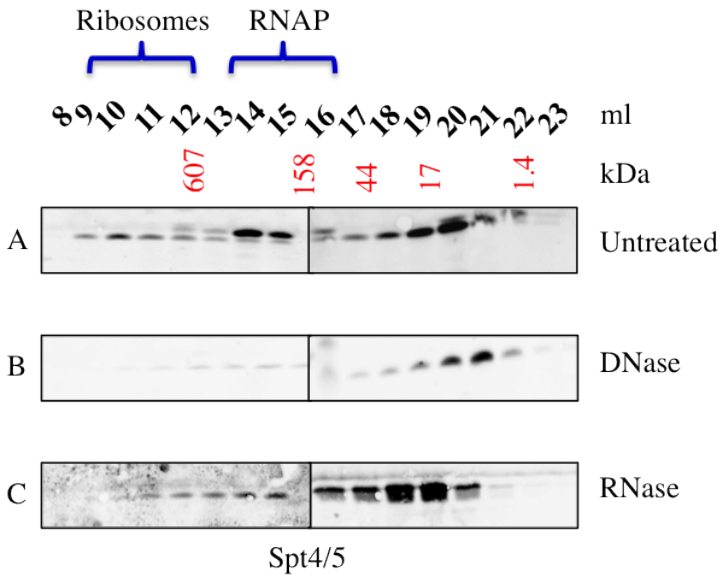


Figure 3.7: Western blot analysis of Superose 6 separated Spt4/5 from cell extract. Comparison with a size standard (red) enabled sizing of complexes. A) In untreated extract Spt4/5 elutes in three peaks corresponding to ribosome-bound, RNAP-bound and free protein. B) Upon DNase treatment of the cell extract Spt4/5 is no longer found in its ribosome or RNAP-bound state. C) RNase treatment destroys the interaction of Spt4/5 with ribosomes and results in a broadening of the elution profile.

3.2.4 Spt4/5-Bound Transcription Complexes are Sensitive to Treatment by Nucleases

After DNase treatment Spt4/5 no longer elutes in fractions containing ribosomes or RNAP (Figure 3.7B). Spt4/5 elutes only in fractions corresponding to free monomeric protein at 20-21 ml. This suggests that the complexes of Spt4/5 with RNAP and ribosomes are DNA dependent. This is consistent with its function as a transcription elongation factor that binds to transcribing RNAP with high affinity.

RNase treatment of the cell extract also abolishes the complex between Spt4/5 and the ribosome (Figure 3.7C) as there is no elution of Spt4/5 in fractions corresponding to a mass in the MDa range. Spt4/5 is still associated with RNAP after RNase treatment, eluting at 14-15 ml. Interestingly the peak corresponding to free Spt4/5 is broader and upshifted, overlapping with

the RNAP-bound Spt4/5. It encompasses fractions from 16-20 ml with its apex at 18-19 ml. This broad elution peak corresponds with a molecular weight ranging from 10-165 kDa.

3.2.5 Analysis of NusA Elution Profile from *M. jannaschii* Extract

NusA elutes in two peaks from the Superose 6 column in untreated cell extract (Figure 3.8A). The first peak is at 11-12 ml, corresponding to a molecular weight in the MDa region. This is approaching the exclusion limit of the column, and it is therefore not possible to accurately determine the mass of the complex. However, the elution profile of NusA overlaps with the elution profile of S10. This therefore raises the possibility that NusA is bound to the ribosome. This is not an unexpected observation because NusA binds to the pre-16S rRNA, promoting its folding and facilitating ribosome biogenesis (Bubunenko et al., 2013). The second elution peak is broad, ranging from 18-21 ml with its apex between 19 and 20 ml. This is free NusA and corresponds to a molecular weight of 14 kDa, which is below the actual NusA mass of 21 kDa. Very little NusA is observed at 13-15 ml, where RNAP elutes.

DNase treatment of the cell extract has little effect on the elution profile of NusA (Figure 3.8B). Given that archaeal NusA consists of two RNA-binding KH domains only and that no RNAP-associated NusA was observed in the untreated cell extract, this is not an unexpected observation.

RNase treatment of the cell extract has a larger effect on the elution profile of NusA but it is still found in fractions containing ribosomes. (Figure 3.8C). As with Spt4/5 and Rpo4/7, the free NusA peak is broader and NusA is present in higher molecular weight complexes compared to extract that has not been treated with RNase. NusA now elutes in one continuous peak from 9-18 ml. This may be due to more free RNA of different sizes being present and bound by NusA, resulting in a broader range of NusA complexes.

Summary

The fractionation of *M. jannaschii* cell extract has demonstrated that Spt4/5 is present in fractions that contain ribosomes and transcription complexes. These complexes are nuclease-

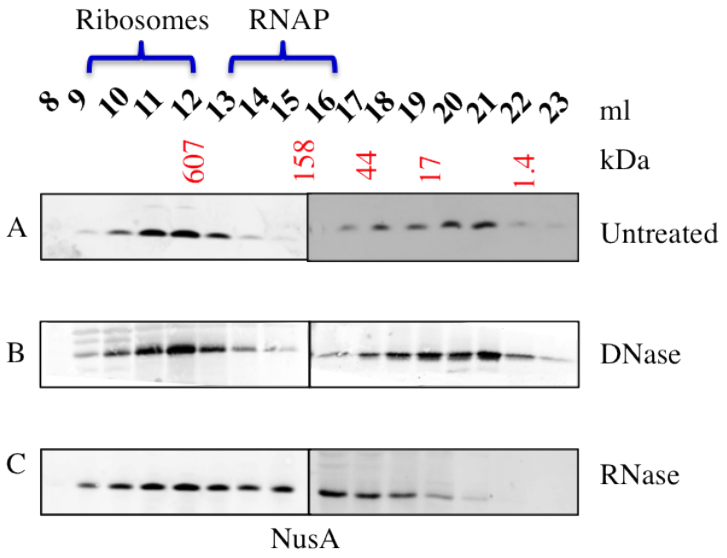


Figure 3.8: Western blot analysis of Superose 6 separated NusA from cell extract. Comparison with a size standard (red) enabled sizing of complexes. A) In untreated extract NusA elutes in two peaks corresponding to ribosome-bound, and free protein. B) DNase treatment of the cell extract has no effect upon the elution profile of NusA. C) RNase treatment results in a broadening of the NusA elution profile.

sensitive, which can be explained by the fact that Spt4/5 is a transcription factor. The presence of RNAP in the ribosomal fractions of untreated extract hints that transcription complexes may be coupled to translation complexes. NusA elutes in fractions containing ribosomes, possibly suggesting a role in translation. The size fractionation provides no evidence of NusA associating with transcription complexes.

3.3 Recombinant Spt4/5 Interacts with Ribosomes

3.3.1 Microscale Thermophoresis

Microscale thermophoresis (MST) is a technique that can be used to quantify interactions between biomolecules. It exploits the fact that a biomolecule has different physical properties when unbound compared to when it is within a complex. These differences are size, hydration shell and

charge. By exploiting multiple differences in molecular behaviour, MST is able to obtain binding data even when small unlabelled molecules bind large labelled molecules. This is particularly useful when dealing with very large molecules that cannot be prepared to high concentrations, such as ribosomes. This is an advantage over techniques such as fluorescence anisotropy which rely on a significant increase in the molecular weight of the labelled complex. In MST one of the interacting proteins is labelled with a fluorescent dye and always kept at constant concentration whilst the other protein is diluted out by serial dilution. MST reactions occur in glass capillaries and require very small volumes. Initially the molecules are evenly distributed throughout the capillary. An infrared laser is then switched on which causes a localised temperature increase, in the μm scale, of about 5°C. The biomolecules within the capillary then migrate along a temperature gradient, away from the source of heat induced by the laser in a phenomenon known as the Soret effect (Duhr and Braun, 2006). The movement along the temperature gradient varies with the aforementioned properties, such that the migrational behaviour of the labelled biomolecule changes as complex formation occurs. At the same time, the fluorescently-labelled molecule is excited by a laser and the fluorescence is monitored. Thus motion of the fluorescently-labelled molecule along the temperature gradient is observable.

There are three parameters that enable binding affinities to be accurately determined. The first is the temperature jump that is induced by the infrared laser. This occurs on a timescale of about 3 seconds. Fluorescence intensity changes with temperature. This can be due to a change in the absorption of the dye, its quantum yield or the lifetime of its fluorescence emission. However, the local environment of the dye also affects these properties. A bound protein can restrict local mobility and can also alter the local chemical environment surrounding the dye. Thus the fluorescence intensity is dependent on both the temperature shift but also on whether a biomolecule is binding the protein near the fluorescent dye (within 2 nm) (Vaiana et al., 2003). The second parameter that enables binding affinities to be determined is thermophoresis in the temperature gradient. This occurs on a slower timescale to the temperature jump, taking about 30 seconds. The final parameter is back diffusion. The thermophoresis results in an inhomogeneous distribution of the fluorescent molecule along the temperature gradient. When the temperature gradient is eliminated the biomolecules will diffuse back to a homogeneous distribution (Figure 3.9). Since there is no longer a temperature gradient present the rate of back diffusion depends only on the

size of the complex. Therefore this phase yields no data when analysing a small unlabelled molecule binding a much larger labelled one (Jerabek-Willemsen et al., 2011).

Figure 3.9: The migration of molecules in an MST experiment. At first the fluorescent molecules are homogeneously distributed throughout the capillary, resulting in a steady fluorescence reading. At approximately 5 seconds the infrared laser is switched on and focussed on the specific region of the capillary that is observed. This creates a temperature gradient. The fluorescently-labelled molecules move down the temperature gradient, resulting in a decrease in fluorescence at the observed point. After approximately 35 seconds the infrared laser is switched off and the temperature gradient dissipates. The fluorescently-labelled molecules diffuse back, resulting in a restoration of the observed fluorescence. Figure from (Jerabek-Willemsen et al., 2011).

3.3.2 Expression of Recombinant Spt4/5

M. jannaschii Spt4/5 was expressed as a GST-fusion protein from a pGEX-2TK vector in Rosetta 2 *E. coli* cells and purified on a glutathione affinity column. The resulting protein was very pure and contained few contaminants (Figure 3.10).

3.3.3 Removal of the GST Tag

The Spt4/5-GST fusion protein was treated with thrombin to separate Spt4/5 from GST. The sample was then incubated at 65°C for 30 minutes, resulting in the precipitation of the GST,

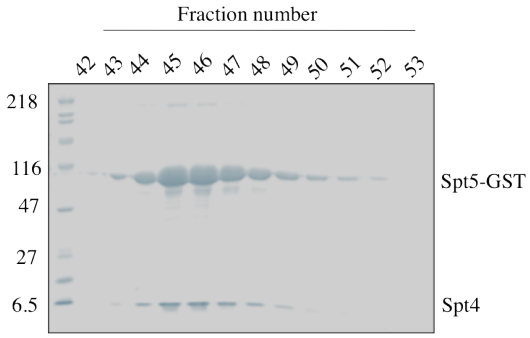


Figure 3.10: SDS PAGE of Spt4/5 purification by GST-affinity chromatography. The GST-tagged protein bound to the column and was eluted by addition of glutathione, yielding a pure GST-tagged protein.

whilst the thermostable Spt4/5 remained folded in solution. The sample was then centrifuged and the resulting supernatant contained the purified Spt4/5 (Figure 3.11).

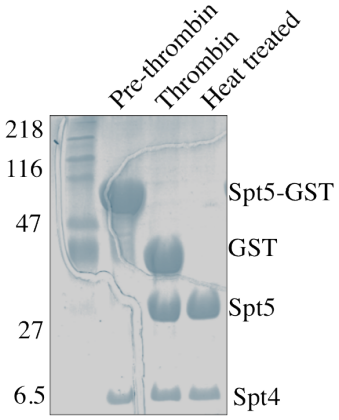


Figure 3.11: SDS PAGE of GST removal and heat-inactivation. The GST tag was removed by overnight treatment with thrombin to give Spt4/5 and GST. The GST was separated from the Spt4/5 by heat inactivation. The sample was heated to 65°C to precipitate the GST out of solution. The sample was centrifuged, pelleting the GST whilst the heat-stable Spt4/5 remained in solution in the supernatant.

3.3.4 Preparation of *M. jannaschii* Ribosomes

Ribosomes were prepared from *M. jannaschii* cell extract (see Chapter 2). The purified ribosomes were analysed on a 10–30% sucrose gradient. The OD_{260} was monitored and fractions

analysed by Western blot against S10 (Figure 3.12A). The RNA from the peak fraction was extracted and was analysed on a bioanalyzer, which revealed the presence of two prominent RNAs, corresponding to the 16S and the 23S rRNAs (Figure 3.12B). The RNA was also analysed on a 1% agarose/TBE gel (Figure 3.12C). [1]

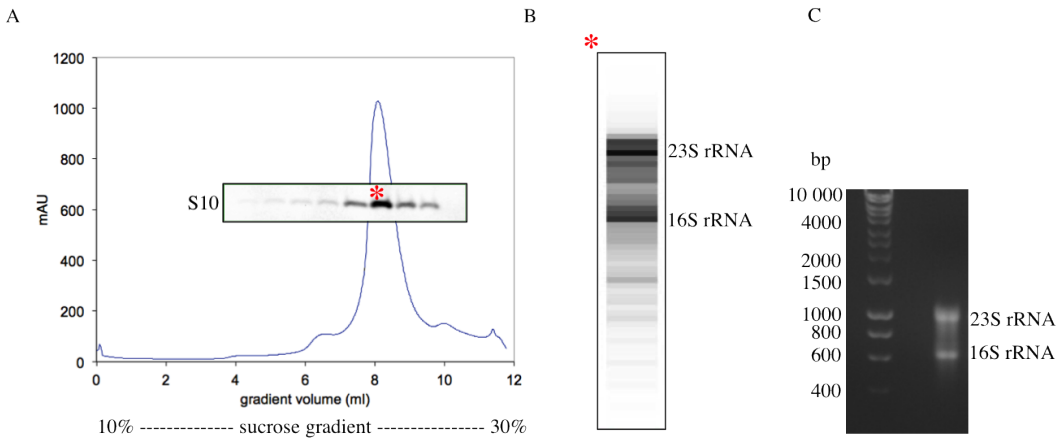


Figure 3.12: Purification of ribosomes from *M. jannaschii* cell extract. A) After washing through a salt cushion to remove bound transcription and translation factors, the ribosomes were analysed on a 10–30% sucrose gradient. The presence of ribosomes was confirmed by analysing the OD_{260}, and by Western blot analysis using antibodies against the ribosomal protein S10. B) The peak fraction (starred) was characterised on a bioanalyzer, revealing the presence of two major RNA bands, corresponding the 23S and 16S rRNAs. C) 1% agarose TBE gel of extracted rRNA. These experiments were conducted by Dr Tina Daviter.

3.3.5 Labelling of Ribosomes with Maleimide-Coupled NT650 Dye

The ribosome preparation was labelled at cysteine residues with NT650. Unbound dye was separated from the labelled ribosomes on a gel filtration drip column, and the labelled ribosomes were eluted in fractions of 100 μl. Fractions containing labelled ribosomes were determined by measuring absorbance at 260 and 650 nm on a Nanodrop UV/Vis spectrometer (Figure 3.13).

3.3.6 Interaction Between Ribosomes and Spt4/5

MST was performed to analyse interactions between recombinant Spt4/5 and ribosomes purified from *M. jannaschii* cell lysate. All experiments were recorded using a Monolith NT.115

[1] Ribosome preparations, bioanalyzer run, and agarose gel electrophoresis were performed by Dr Tina Daviter.

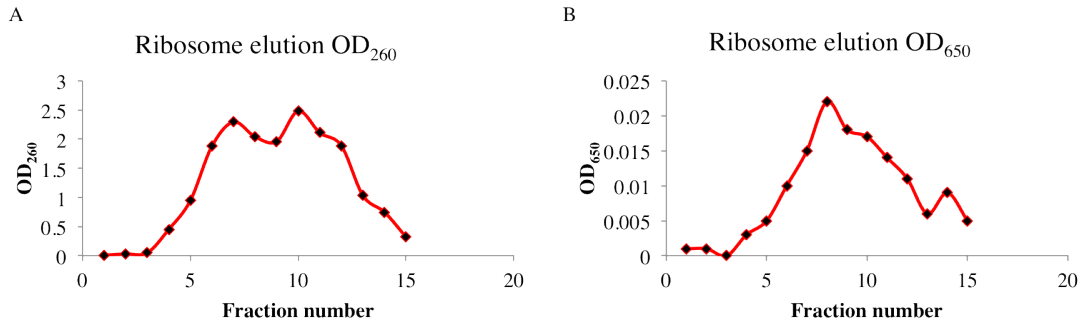


Figure 3.13: Elution profiles of labelled ribosomes. Labelled ribosomes were separated from free label on a gel filtration column. A) OD_{260} corresponding to ribosomes. B) OD_{650} corresponding to labelled ribosomes.

instrument. A serial dilution of Spt4/5 from 3 nM to 100 μM was set up. Ribosomes at a final concentration of 15 nM, labelled with a maleimide-coupled dye NT650 were incubated with Spt4/5 at room temperature for 30 minutes. The samples were then excited at 650 nm, and the infrared laser turned on 5 seconds later. Thermophoresis was analysed for 30 seconds, then the infrared laser was turned off and back diffusion was analysed for 5 seconds thereafter. Plotting the temperature jump against the concentration of Spt4/5 yields a binding curve, suggesting that Spt4/5 is interacting with the purified ribosomes (Figure 3.14A). From this binding curve a K_d of 2 μM was obtained for the interaction. This is the first time that an interaction between Spt4/5 and whole ribosomes has been observed. A negative control whereby TATA-binding protein (TBP) was titrated into the labelled ribosomes did not result in binding (Figure 3.14B), thereby demonstrating that the decrease in fluorescence upon addition of Spt4/5 is due to specific binding.

3.3.7 The KOW Domain of *M. jannaschii* Spt4/5 Interacts with the Ribosome

Previous studies in *E. coli* have demonstrated that only the KOW domain of NusG interacts with the ribosomal protein S10. To determine whether it is the KOW domain that binds ribosomes in archaea too and whether the NGN domain is involved in the interaction in the context of complete ribosomes, further MST analysis was performed on two domain-deletion variants of Spt4/5. The first was Spt4/5 NGN which consists of the NGN domain of Spt5 in complex with Spt4 but

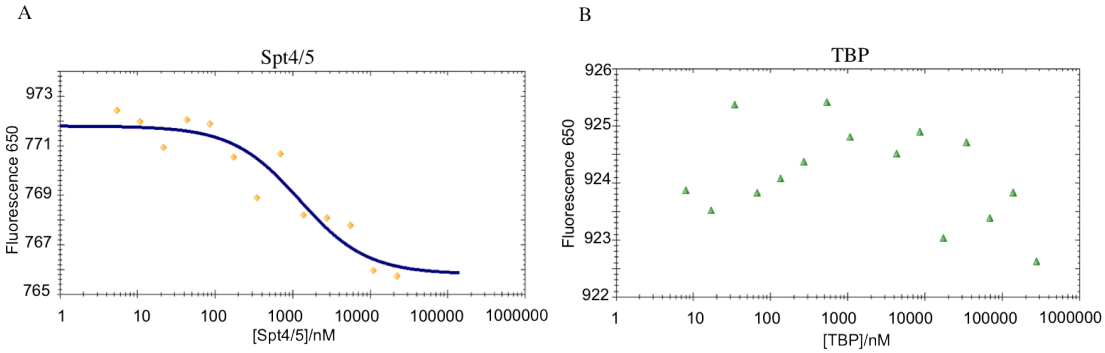


Figure 3.14: MST analysis indicates an interaction between recombinant Spt4/5 and ribosomes prepared from *M. jannaschii* cell extract. A) Titration of Spt4/5 into labelled ribosomes. B) Titration of TBP into labelled ribosomes. Data obtained at NanoTemper MST demonstration with Dr James Wilkinson.

which lacks the KOW domain. The second was the KOW domain only, lacking both the NGN domain of Spt5 and Spt4. The experiments were carried out in the same way as for full length Spt4/5. When MST was conducted using labelled ribosomes and the KOW domain, binding was observed and analysis of the data yields a K_d of 4 μM (Figure 3.15A). This is in good agreement with full length Spt4/5. When MST was carried out using labelled ribosomes and Spt4/5 NGN, no binding was observed (Figure 3.15B) indicating that the NGN domain is not involved and that Spt4/5 interacts with ribosomes solely through the KOW domain as it does in bacteria.

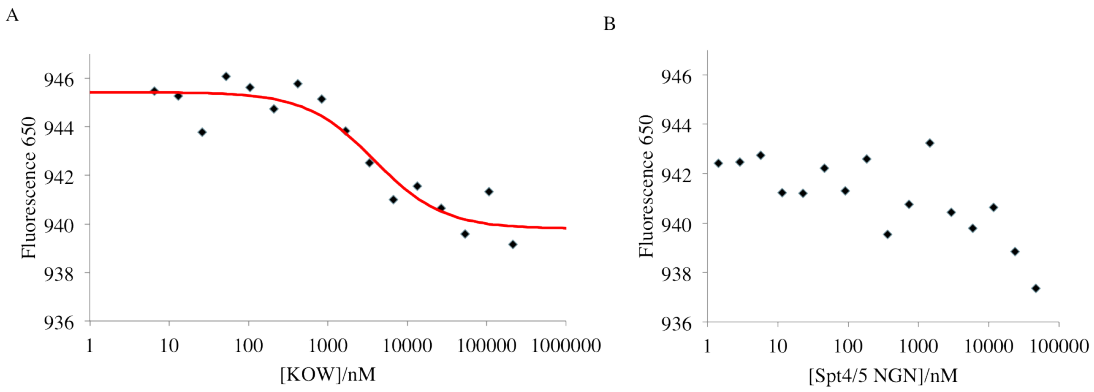


Figure 3.15: MST analysis using labelled ribosomes prepared from *M. jannaschii* cell extract and recombinant domain deletion variants of Spt4/5. A) The KOW domain interacts with ribosomes. B) Spt4/5 NGN does not interact with ribosomes.

The interaction between the KOW domain and ribosomes was further characterised by pre-

incubating the labelled ribosomes with either anti-S10 or anti-TBP antibodies that had been purified from rabbit serum (Figure 3.16). When the ribosomes were incubated with anti-S10 no binding was observed upon addition of the KOW domain (Figure 3.17A). This indicates that the anti-S10 antibodies were blocking the binding site and that the KOW domain must therefore bind to the S10 protein. When the labelled ribosomes were incubated with anti-TBP there was no change in the MST analysis, and the KOW domain was shown to bind the ribosomes with a K_d of 3 μM, indicating that the effect of the anti-S10 is not due to non-specific binding of the antibodies (Figure 3.17B).

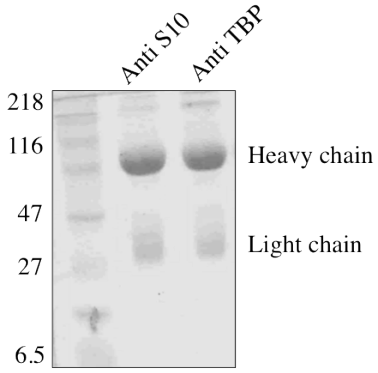


Figure 3.16: SDS PAGE of purified polyclonal antibodies. The heavy and light chains are indicated.

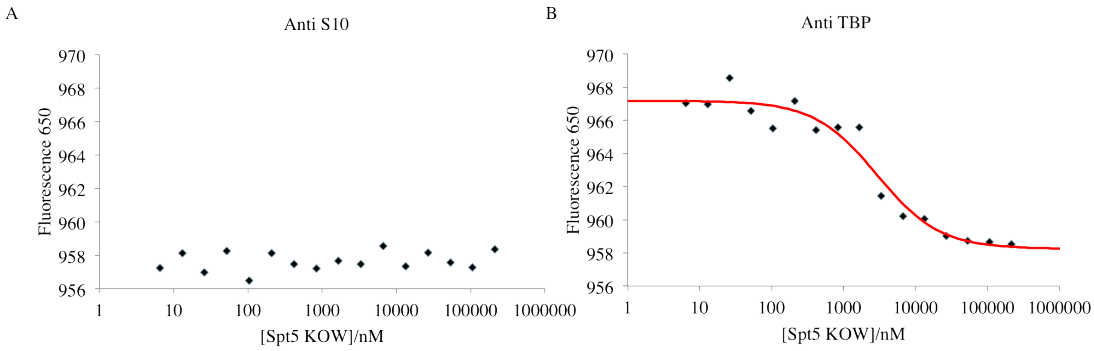


Figure 3.17: MST analysis was performed on labelled ribosomes and the KOW domain in the presence of polyclonal antibodies. A) When incubated with anti-S10 no binding is observed indicating that anti-S10 abolishes the interaction. B) When incubated with anti-TBP the binding face on the S10 subunit is still available for binding to the KOW domain.

3.4 Interactions Between Spt4/5 and Recombinant S10

Having observed interactions between *M. jannaschii* Spt4/5 and ribosomes, experiments were performed to attempt to observe interactions between purified recombinant factors. To this end Superose 12 size exclusion chromatography on Spt4/5 and S10 was carried out. First Spt4/5 and S10 were separated on the column individually and their elution volume determined by SDS PAGE and silver staining. Spt4/5 and S10 were subsequently incubated together at 65°C for 30 minutes and the pre-incubated sample was separated on the Superose 12 column. Further SDS PAGE and silver staining was conducted on the pre-incubated sample.

3.4.1 Purification of Recombinant S10

The ribosomal protein S10 was expressed in Rosetta 2 cells from a pET-21a(+) vector with no affinity tag. After cell extraction, the soluble portion of the cell extract was heated-inactivated to precipitate most of the endogenous *E. coli* proteins. The heat-inactivated sample was then centrifuged to separate the soluble and insoluble portions. The protein in the soluble portion was precipitated by adding ammonium sulphate. The precipitate was pelleted, resuspended in 5 ml N500 buffer and purified by size exclusion chromatography on an S100 column (Figure 3.18). The fractions containing pure protein were then concentrated.

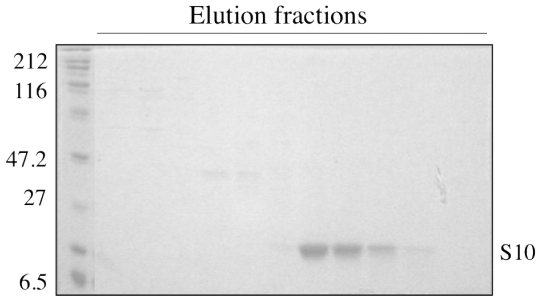


Figure 3.18: SDS PAGE of S10 purification on a S100 column.

3.4.2 Superose 12 Size Exclusion Chromatography on Spt4/5 and S10

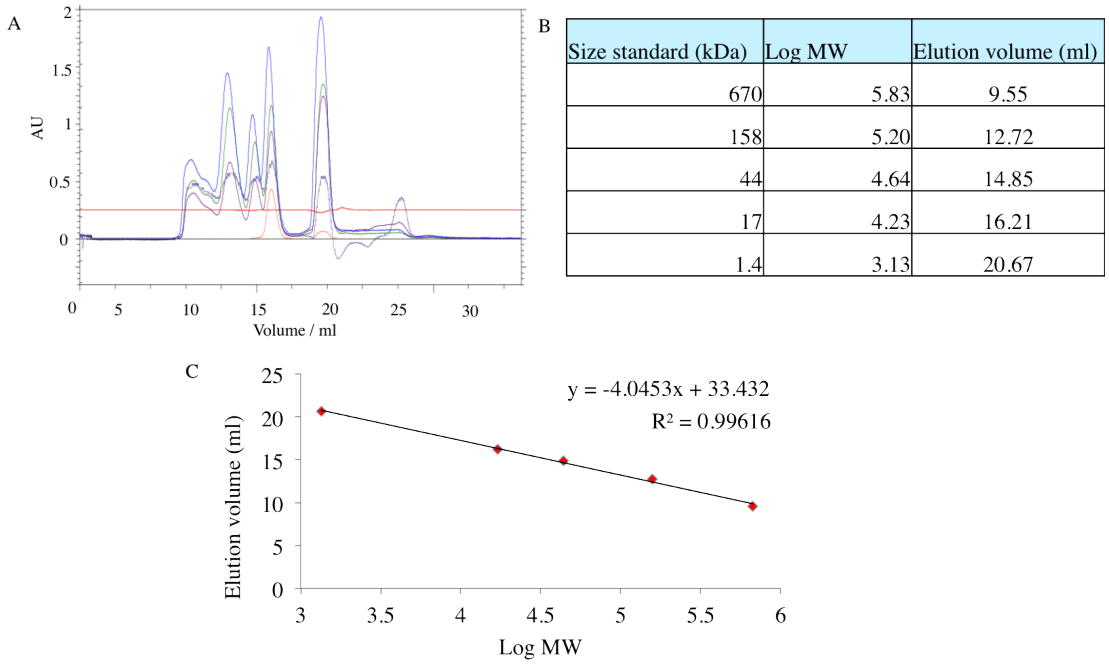
First, a size standard was run on the Superose 12 column to enable sizing of any proteins or complexes (Figure 3.19). Plotting the log of the molecular weight against the elution volume enabled the generation of a standard curve (Figure 3.19C) with the equation:

$$y = -4.04x + 33.4 \tag{3.3}$$

where y is the elution volume and x is log MW. This can be rearranged to:

$$x = (33.4 - y)/4.04 \tag{3.4}$$

so that for any given elution volume, the molecular weight can be calculated.



Size standard (kDa)	Log MW	Elution volume (ml)
670	5.83	9.55
158	5.20	12.72
44	4.64	14.85
17	4.23	16.21
1.4	3.13	20.67

Figure 3.19: A) Elution volumes of the five proteins in the size standard. B) Plotting the elution volume of the proteins in the size standard against log MW enables the generation of a standard curve. Using the standard curve the molecular weight of any complex can be determined, given its elution volume.

When separated on the column on its own Spt4/5 elutes in a broad peak with the apex at 16 ml

(Figure 3.20A). Comparison of this volume with a size standard indicates a molecular weight in the region of 20.3 kDa which is consistent with Spt4/5 existing as a monomer in solution. When pre-incubated with S10, the elution profile of Spt4/5 is the same, showing a broad distribution with the apex of the peak at 16 ml (Figure 3.20C). Again this is consistent with a monomeric Spt4/5 and indicates that under the conditions present in the Superose column, there is no stable interaction occurring between Spt4/5 and S10.

The elution profile of S10 supports this conclusion. When separated on the column by itself S10 elutes at 16-17 ml (Figure 3.20B). By comparison with the size standard, 17 ml corresponds to a molecular weight of 11.5 kDa. This is consistent with a monomeric S10 which has an actual MW of 12.2 kDa. When preincubated with Spt4/5, S10 still elutes at 16-17 ml (Figure 3.20C). The absence of a decrease in elution volume upon preincubation of the proteins indicates that no stable complex has been formed between Spt4/5 and S10.

3.5 NusA is Stably Associated with the Ribosome

The size fractionation of cell extract showed that NusA elutes in the same fractions as ribosomes. The presence of NusA in ribosomes was confirmed by preparing ribosomes that had been washed in 500 mM NH_4Cl to remove any loosely-associated translation factors. The ribosomes were then separated on a Superose 12 column and the fractions analysed by Western blot using polyclonal antibodies against S10 and NusA. The ribosomes elute from the Superose 12 column as one peak with a slight shoulder at the higher molecular weight end. This indicates that the ribosomes exist largely as one species, and have not separated into their constituent large and small subunits. The peak maximum is between 8-9 ml (Figure 3.21A). S10 elutes in one peak at 8-9 ml, in good agreement with the elution OD_{260} profile, and acting as a positive control for the presence of ribosomes (Figure 3.21B). NusA is present in the same fractions as S10, indicating that it is stably associated with the ribosome (Figure 3.21C).

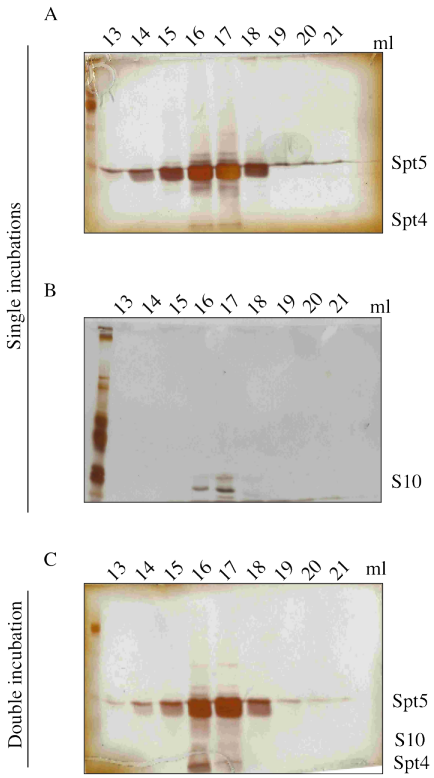


Figure 3.20: Silver stain analysis of Superose 12 size exclusion chromatography of Spt4/5 and S10. Elution volumes in ml are indicated above the lanes. A) Spt4/5 separated by size exclusion chromatography followed by SDS PAGE and silver staining. B) S10 separated by size exclusion chromatography followed by SDS PAGE and silver staining. C) Spt4/5 preincubated with S10 and separated by size exclusion chromatography followed by SDS PAGE and silver staining. Incubating Spt4/5 and S10 together does not result in a shift in either of their elution profiles, suggesting that no stable interaction is formed.

3.6 Discussion

In this chapter, size exclusion chromatography of *M. jannaschii* cell extract followed by Western blot analysis has shown that Spt4/5 is in fractions that contain i) RNAP and ii) ribosomes. The nature of these complexes has been verified by using antibodies against the recombinant RNAP subunits Rpo4/7 and the ribosomal protein S10. This suggests that Spt4/5 is present in both transcription and translation complexes. This chapter has also shown that the complex between Spt4/5 and RNAP is dependent upon DNA, since DNase treatment of the cell extract abolishes the coelution of RNAP and Spt4/5. This is consistent with previous reports that Spt4/5 binds to transcription elongation complexes (Grohmann et al., 2011). The presence of RNAP sub-

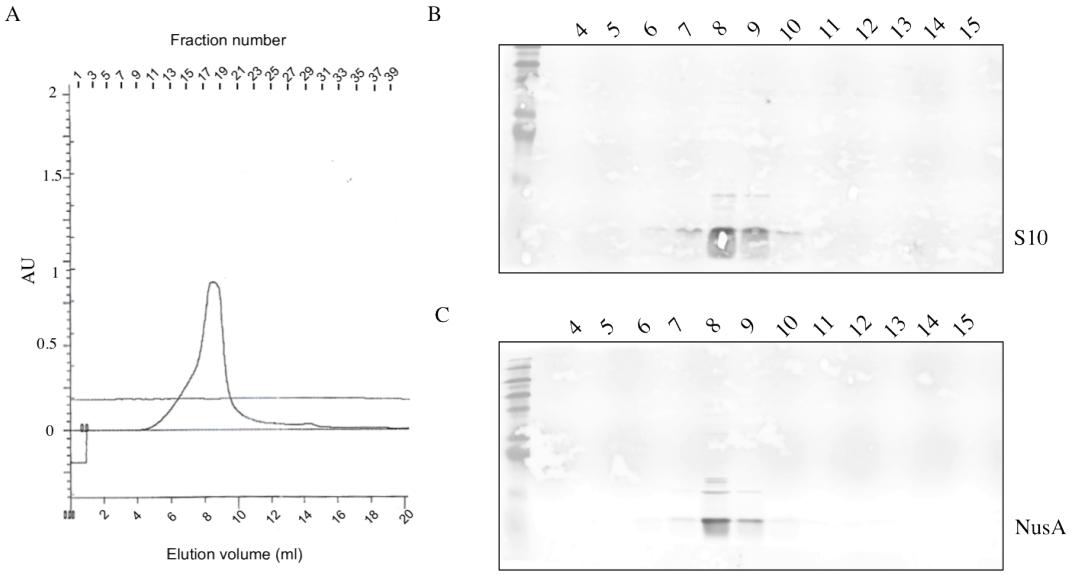


Figure 3.21: Association of NusA with the ribosome. A) Purified ribosomes were separated on a Superose 12 size exclusion column and the absorbance at 260 nm was monitored. Western blots were carried out on the separated ribosome fractions to test for the presence of B) S10, and C) NusA. S10 and NusA both elute at 8–9 ml, indicating that NusA is stably associated with the ribosome.

units Rpo4/7 in ribosomal fractions raises the possibility that Spt4/5 binds to the ribosome and RNAP simultaneously. This could be further examined *in vivo* by knocking out wild-type Rpo1′ and replacing it with the well-characterised coiled-coil deletion mutant in which the coiled-coil residues (255–264) are replaced with a tetra-glycine linker (Hirtreiter et al., 2010a). This mutant is unable to interact with Spt4/5. If RNAP were no longer found in the ribosomal fraction after introduction of this mutation, it would provide strong evidence that Spt4/5 is coupling RNAP and the ribosome. This experiment could not be performed in *M. jannaschii* as it is not genetically tractable but could be conducted in other archaeal species such as *S. acidocaldarius*.

NusA was found to elute in two peaks corresponding to ribosome-associated and free protein. The association of NusA in the ribosomes was confirmed by performing Western blots on purified ribosomes that had been separated on a size exclusion column. The lack of any NusA in fractions corresponding to RNAP suggests that NusA does not bind to RNAP.

RNase treatment of the cell extract results in a broadening of the elution profile on the lower molecular weight regions for Spt4/5, Rpo4/7 and NusA. For Rpo4/7 and NusA which are known

to bind to RNA it is possible that RNase treatment results in the cleavage of large unavailable RNAs in to smaller fragments of differing sizes. These liberated RNAs are then free to interact with NusA and Rpo4/7, resulting in them eluting earlier from the column due to increased molecular weight. Although archaeal Spt4/5 is not reported to bind nucleic acids in the literature, it has become apparent through development of purification strategies that it copurifies with large amounts of nucleic acids which can be removed by heparin affinity purification. Therefore this hypothesis can also explain why the Spt4/5 peak broadens upon RNase treatment.

MST was conducted to quantitatively assess the interaction between Spt4/5 and the ribosome. It was shown that Spt4/5 interacts specifically with the ribosome via the KOW domain. It was demonstrated that Spt4/5 interacts with the S10 subunit of the ribosome since preincubation with polyclonal antibodies against S10 abolishes the interaction. MST gave a K_d for the interaction in the region of 2-4 μM. This affinity is consistent with gene expression being a dynamic process. The interaction must be strong enough to enable the robust coupling of transcription and translation. Yet the interaction must also not be too tight. After the RNAP has terminated transcription it needs to be able to reinitiate transcription. Similarly the ribosome needs to reinitiate translation. This would not be possible if they were permanently coupled through Spt4/5 with low nM affinity.

Size exclusion chromatography was then carried out between Spt4/5 and recombinant S10. However, specific interactions were not detected. This suggests that within the complexity of the ribosome there is an increased affinity for Spt4/5 and that when the ribosomal subunit S10 is isolated its affinity for Spt4/5 is decreased.

Another method that could be used to confirm the interactions observed by MST is co-pelleting. In co-pelleting assays ribosomes are washed with salt to remove any loosely bound factors and then incubated with the protein of interest. The incubation is centrifuged at high speed to pellet the ribosomes which are then analysed by Western blot for the protein of interest. If the protein interacts with the ribosome, it will be present in the pellet and give a positive signal in the blot. This assay has the advantage of not requiring labelling of the ribosome, which may interfere with binding events. The interactions between Spt4/5 and the ribosome could be further

characterised by conducting site-directed mutagenesis of the Spt5 residues predicted to interact with S10. In *E. coli*, for example, F165 of NusG is very important for the interaction with S10 (Burmann et al., 2010b).

The biological significance of the interaction could then be determined *in vivo*. For example, what would be the effect on gene expression of abolishing the Spt4/5-ribosome interaction? Would there be decreased transcription due to polarity? This could be examined by developing an assay for determining *in vivo* transcription and translation rates, in a similar manner to that which has been reported in bacterial systems (Proshkin et al., 2010; Vogel and Jensen, 1994). Subsequent mutations could be introduced into the Spt5 KOW domain to prevent it binding the ribosome. If this mutation disrupted the ratio between transcription and translation rates, it would provide very strong evidence that Spt4/5 couples transcription and translation.

Chapter 4

Biophysical Analysis of Spt4/5

4.1 Introduction

4.1.1 The Role of NusG Proteins in the Three Domains of Life

NusG and its homologue Spt5 are the only known universally conserved RNAP-associated transcription factors, suggesting an ancient role in transcriptional regulation. NusG and Spt5 exhibit strong sequence and structural homology across the three domains of life (Figure 4.1), which demonstrates their key role in the regulation of gene expression and suggests that NusG-like proteins were present in the last universal common ancestor. The NGN domain has been shown to bind to RNAP in all three domains of life (Hirtreiter et al., 2010b; Mooney et al., 2009b; Wenzel et al., 2010), enclosing the DNA binding channel and thereby increasing the processivity of RNAP (Burova et al., 1995; Hirtreiter et al., 2010b; Wada et al., 1998).

Despite the conservation of structure and function, NusG proteins have also evolved additional functions. The NGN domain of archaeal and eukaryotic Spt5 forms a stable complex with Spt4 but there is no Spt4 homologue in bacteria (Guo et al., 2008). One important question to be addressed, therefore, is the function of Spt4 in archaea. Early studies indicated that Spt4 increases the stability of the Spt5 NGN domain (Hirtreiter et al., 2010a), however this has yet to be quantified. Furthermore, whilst the NGN domain of Spt5 is necessary to stimulate transcription elongation in archaea, it is not sufficient and requires either Spt4 or the Spt5 KOW domain (Hirtreiter et al., 2010a). This is in contrast to bacteria where the NusG NGN domain alone is sufficient to stimulate RNAP elongation (Mooney et al., 2009b). The function of the KOW domain also shows variation across the three domains of life. In bacteria and archaea the KOW domain is not required for recruitment of NusG/Spt5 to RNAP. However, evidence from eukaryotes suggests that the KOW domains show weak affinity for RNAP (Viktorovskaya et al., 2011) and that the KOW domains bind to RNA, possibly as a requirement for recruitment to RNAP (Cheng and Price, 2008). In addition, the KOW domain is believed to bind to the ribosome in bacteria and archaea, thereby coupling transcription and translation (Burmann et al., 2010b). In eukaryotes, however, the processes of transcription and translation occur in different cellular compartments and cannot be directly coupled. Consequently the KOW domain has evolved alternative functions in eukaryotes, acting as a recruitment platform for proteins involved in RNA processing (Lindstrom et al., 2003) and promoter-proximal stalling (Missra and Gilmour, 2010).

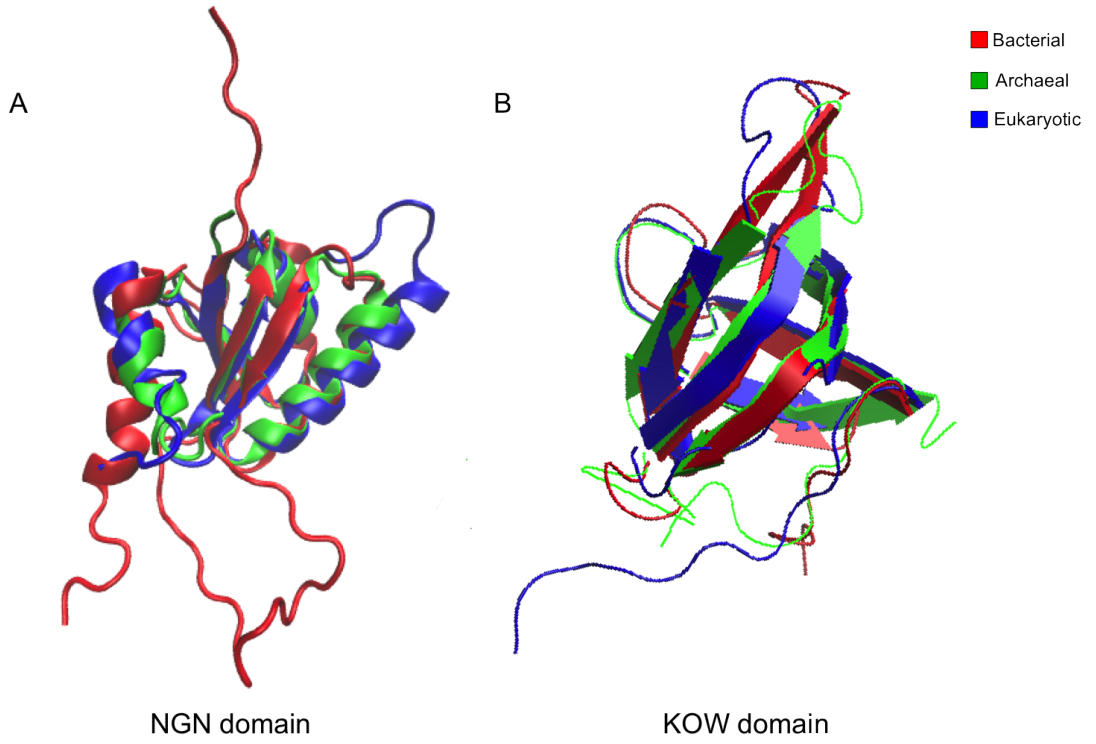


Figure 4.1: Structural alignment of the NusG/Spt5 domains from the three domains of life. A) The NGN domain exhibits high structural conservation, which is indicative of its fundamental role in the regulation of gene expression. Red: *E.coli* NusG NGN domain (PDB 2KO6). Green: *M. jannaschii* Spt5 NGN domain (PDB 3LPE). Blue: *S. cerevisiae* Spt5 NGN domain (PDB 2EXU). B) The KOW domains also overlay, comprising a five-stranded β barrel. Red: *E.coli* NusG KOW domain (PDB 2JVV). Green: *P. furiosus* Spt5 KOW domain (PDB 3P8B). Blue: *H. sapiens* Spt5 NGN KOW 2 domain (PDB 2E6Z).

4.1.2 Chapter Aims

The aims of this chapter are to discover the contributions of each domain to the structure and integrity of the Spt4/5 complex. To this end pulsed electron paramagnetic resonance (EPR) techniques were performed on double spin-labelled Spt5 in the absence and presence of Spt4 to ascertain whether Spt4 influences the topology of the two Spt5 domains. Domain-deletion mutants of Spt4/5 were also analysed by a thermal denaturation assay and by limited proteolysis. The aims of these experiments were to determine which domains contribute to the stability of

Spt4/5.

4.2 Electron Paramagnetic Resonance

The most common technique for determining protein structure is X-ray crystallography with over 90% of structures in the protein data bank being determined by this method (Berman et al., 2000). The biggest advantage of X-ray crystallography is the ability to resolve structures to within a few Å. However in order to achieve good scattering the protein must be arranged in an orderly crystal and be immobile. Proteins however, are not static entities but dynamic molecules which are often capable of undergoing large conformational changes. It is therefore difficult to ascertain whether a structure obtained by crystallography represents a biologically relevant state or that of an inactive conformation. Furthermore different crystal structures of the same protein may present drastically different conformations. This issue can be circumvented by using solution methods such as nuclear magnetic resonance (NMR). NMR is able to give structures of proteins in solution. However protein concentrations in the region of 400 μM are required, which is not always possible. The protein preparation is costly to produce, requiring isotopic labelling with ^{15}N, ^{13}C, and ^{2}H. Furthermore as molecules increase in mass, their slower tumbling results in individual resonances broadening and overlapping, making assignment very difficult. Consequently NMR is not routinely performed on proteins over 35 kDa.

An alternative technique to obtain structural information about a protein is electron paramagnetic resonance (EPR), which analyses unpaired electrons. A protein molecule containing an unpaired electron is said to be paramegnetic. Although most proteins do not contain paramagnetic centres, it is possible to incorporate an unpaired electron into a protein by site-directed spin labelling (SDSL) (Altenbach et al., 1989, 1990; Hubbell and Altenbach, 1994). This is achieved by mutating residues of interest to cysteine and expressing the mutant, which will then contain a cysteine in place of the wild-type residue at the selected position. The newly-introduced cysteine can then be reacted with a nitroxide spin label that exhibits selectivity for sulfhydryl groups. The two most commonly used cysteine-specific nitroxide spin labels are methanethiosilfonate spin label (MTSSL) and 3-(2-Iodoacetamido)-PROXYL. Site-directed spin labelling may also require replacing any native cysteines already present in the protein, usually with a serine or an alanine (Klare and Steinhoff, 2009). The end result is a protein which is specifically labelled

at the newly-introduced cysteine residue or residues, but nowhere else. It is also possible to incorporate spin labels in a site-specific manner into DNA (Edwards and Sigurdsson, 2007; Ward et al., 2007) and RNA (Edwards et al., 2001), which is potentially very useful in the study of transcription.

4.2.1 Continuous Wave EPR

An electron can exist in two spin states, -1/2 or +1/2. In the absence of an external magnetic field the energies of the two states are degenerate. However, if an external magnetic field is applied the energies of the two spin states separate, in a phenomenon known as the Zeeman effect (Figure 4.2). The greater the strength of the external magnetic field, the greater the energy difference between the two spin states. Spins that are aligned parallel to the magnetic field have a lower energy, and spins that are aligned antiparallel to the magnetic field have a higher energy. An electron in the lower energy level can be excited into the higher energy level by the absorbance of microwaves whose frequency is equal in energy to that between the two electron spin states. The absorbance of energy, and excitation of an unpaired electron from a ground to an excited state forms the basis for all EPR experiments. As the magnetic field strength that the electron is placed within increases, the frequency of microwave energy required to achieve resonance increases proportionally. In an X-band EPR experiment a microwave frequency of approximately 9.5 GHz is used, corresponding to a magnetic field in the region of 3300 G.

In a continuous wave (CW) EPR experiment, the sample is placed within an external magnetic field and irradiated with microwave energy of constant frequency. The magnetic field is then swept and the first derivative of the absorbance spectrum recorded. The CW EPR spectrum of a nitroxide spin label appears as three peaks (Figure 4.3). This is due to the phenomenon of hyperfine coupling whereby atoms experience a small local magnetic field from neighbouring atomic nuclei. The nitrogen atom that is bonded to the paramagnetic oxygen within a nitroxide spin label has a spin number of 1, and can therefore exist in three spin states: -1, 0, or +1, each with a different energy level. This results in the paramagnetic oxygen experiencing one of three potential local magnetic fields, each local magnetic field corresponding to one of the three possible spin states of the nitrogen. Consequently, for a given microwave frequency, there are

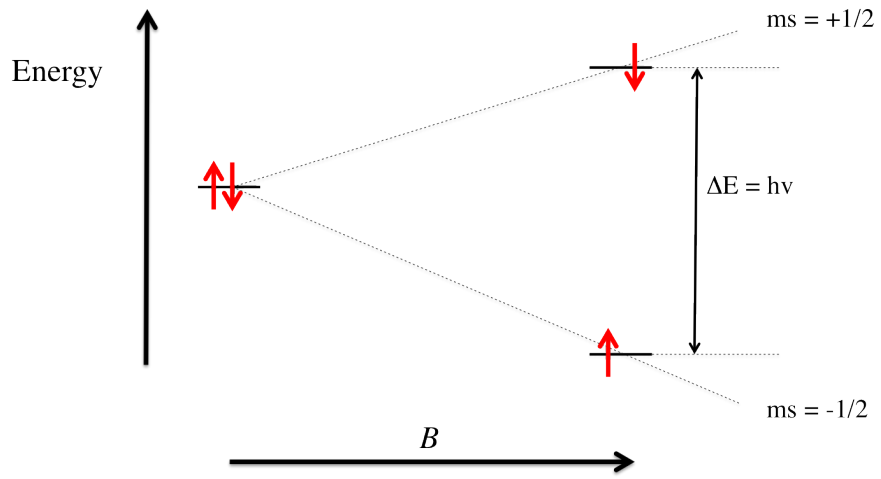


Figure 4.2: Zeeman splitting of an electron in a magnetic field. An electron can exist in two possible spin states. When placed in an external magnetic field (designated B) the two spin states differ in energy. The difference in energy between the two spin states increases with the magnetic field. This phenomenon is called Zeeman splitting.

three different magnetic field strengths that can enable the transition of a nitroxide spin label to its excited state within a sample.

The shape of a CW EPR spectrum can yield much information about local structural elements of a protein, such as secondary structure (Hubbell et al., 1998, 2000). Spin labels coupled to cysteines in loops typically exhibit high motional freedom, resulting in a sharp spectrum. The motion of spin labels coupled to cysteines on α helices and β sheets is more restricted, and results in a broader spectrum. CW EPR can therefore be used to ascertain whether a spin label has successfully been coupled to a protein, since the spectrum of a nitroxide spin label is much broader when bound to a protein than when free in solution (Figure 4.4).

4.2.2 Pulsed EPR

When two unpaired electrons are in close proximity, the energy of the first may influence the energy of the second, in a phenomenon known as dipolar coupling. This means that the spin of one electron will influence the frequency at which its neighbour absorbs microwave radiation. If spin A is aligned parallel to spin B, then spin B will be in a less energetically favourable position

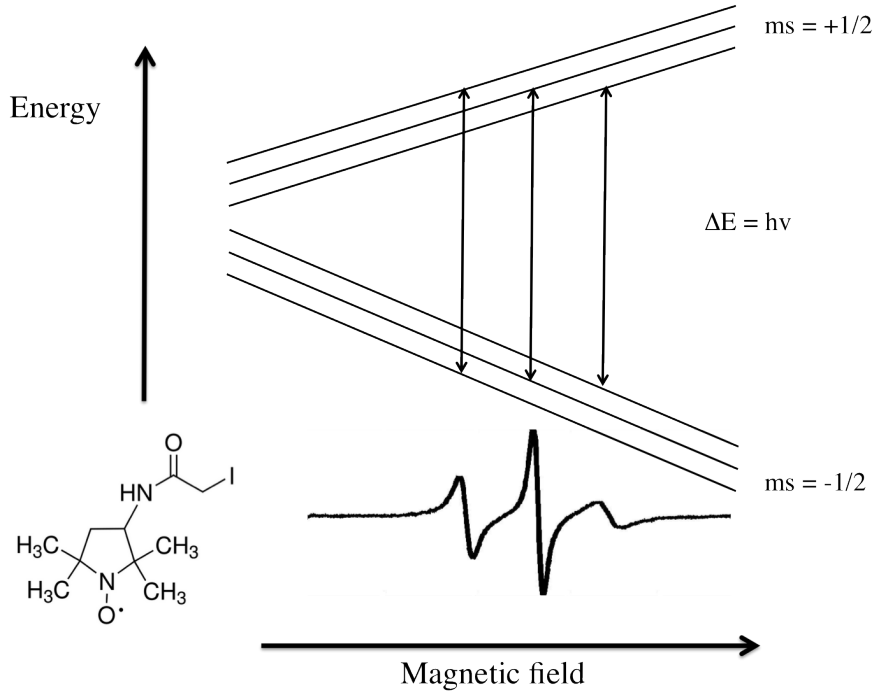


Figure 4.3: A CW EPR spectrum of a nitroxide spin label such as 3-(2-Iodoacetamido)-PROXYL (bottom left) appears as three peaks due to a phenomenon called hyperfine coupling. The paramagnetic oxygen of the label is bonded to a nitrogen which itself exists in one of three possible spin states, each with a different energy. Therefore, for a given microwave frequency there are three magnetic field strengths at which resonance can be achieved.

Figure 4.4: CW EPR spectrum showing different spin label mobilities. The more motion a spin label can undergo, the narrower the spectral peaks.

and will have a higher ground-state energy. It will therefore require microwaves of lower energy to be excited into its excited state. If spin A is aligned antiparallel to spin B, then spin B will be in a more energetically favourable position and will have a lower ground-state energy. It will therefore require microwaves of higher energy to flip into its excited state. The strength of the dipolar interaction between the two electrons is proportional to $1/r^3$. Thus by monitoring the strength of dipolar interaction between two paramagnetic centres, it is possible to determine distance between them. The experiment used to determine the distance between two spin labels

is called double electron-electron resonance (DEER) (Jeschke and Polyhach, 2007) and can normally measure distances up to 8 nm (Schiemann and Prisner, 2007).

In order to describe the DEER experiment it is necessary to define the spin label axis. The x-axis runs parallel to the N-O bond of the nitroxide. The z-axis is perpendicular to the plane of the ring of the label. The y-axis is perpendicular to both the x and z-axis. At equilibrium the external magnetic field, B_0, is aligned with the z-axis. The spins will be aligned along the external magnetic field, in one of two spin states, which are either parallel or antiparallel to B_0. Since it is more energetically favourable to be aligned parallel, there is bulk magnetisation in the $+x$-axis. The spins can be flipped and focussed into the x-y plane by providing a microwave pulse of specific length. This is called a $\pi/2$ pulse. The focussed spins in the x-y plane constitute a divergence from equilibrium and relaxation occurs to restore it. Relaxation occurs in two ways.

Spin Relaxation

T1, or spin-lattice relaxation, describes the loss of energy from the system as it returns to equilibrium. This is how quickly the spins realign with the external magnetic field along the z-axis. T1 determines the repetition rate of the pulses. T2, or spin-spin relaxation describes how quickly the aligned spins in the x-y plane 'fan out', or lose their coherence. T2 determines how long data can be collected for an experiment, and therefore the length of distances that it is possible to measure. It is possible to refocus the dispersing spins within the x-y plane by providing a 180°, or π pulse at time t after the initial $\pi/2$ pulse. This flips the dispersing spins 180°, resulting in them precessing in the opposite direction around the x-y plane and refocussing at time t after the second pulse.

DEER works on the principle that it is possible to selectively excite different populations of electrons within a system by using two different microwave frequencies (Figure 4.5). One set of electrons called the observer electrons are excited with a $\pi/2$ pulse to causing the magnetisation precess around the x-y plane. These electrons become dispersed as the spins precess at slightly different frequencies due to spin-spin relaxation. The same set of observer electrons are then excited with a π pulse to refocus the magnetisation, resulting in a Hahn echo. The magnetisa-

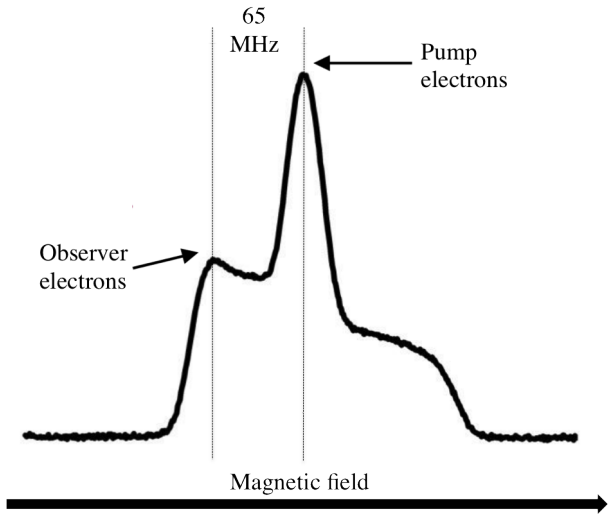


Figure 4.5: Example of an EDFS revealing discrete populations of electrons. The observer and pump electrons are excited separately in the DEER experiment and are separated by 65 MHz. The effect of exciting the pump electrons is monitored by the ability of the observer electrons to refocus.

tion is allowed disperse again and a second π pulse is applied to the observer electrons and the resulting echo recorded. However, between the first echo and second π pulse on the observer electrons, a second set of electrons, called the pump electrons, are excited with a π pulse. Since the pump electrons undergo dipolar interactions with the observer electrons, exciting the pump electrons will alter the local energy of the observer electrons, thereby affecting their ability to be refocussed, and resulting in a decrease of the recorded echo intensity. The series of pulses is repeated, but the time at which the pump electron are excited is shifted (Figure 4.6). As this time increases, the ability to refocus the magnetisation of the observer electrons decreases further and thus the detected echo further decreases in intensity. This can be used to determine the precise distance between the two dipolar-coupled electrons.

4.2.3 Determining T2 of the Sample

Prior to performing the DEER experiment, it is necessary to determine the spin-spin relaxation time, or T2. T2, determines the length of time that signal can be recorded. The longer the T2, the greater the possible distance that can be measured. T2 can be extended by maintaining the sample in D_2O-based buffer. By also deuterating the protein it is possible to measure distances

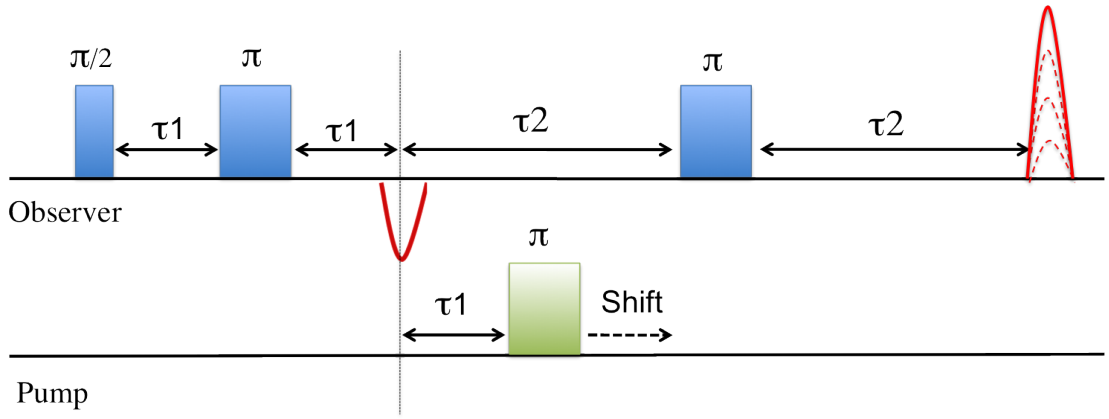


Figure 4.6: Schematic of a DEER experiment. One subset of spins (the observer spins) is flipped with a $\pi/2$ pulse to focus them. After a time, $\tau 1$, the observer spins which have started to dephase, are refocussed with a $\pi/2$ pulse resulting in a Hahn echo. After a second time, $\tau 2$, the observer electrons are refocussed again and the recording echo signal recorded. Between the first echo and the second π pulse on the observer electrons, a second subset of electrons (the pump electrons) are excited. Due to dipolar coupling, this reduces the ability of the observer electrons to refocus. The time at which the pump electrons are flipped shifts with each round of the experiment, further limiting the extent to which the observer electrons can be refocussed, and thereby decreasing the recorded signal.

of up to 10 nm (Ward et al., 2010). DEER experiments are typically performed at 50 K, which is a trade-off between having a long T2 (and therefore the ability to measure longer distances) and a shorter T1 (thereby increasing the repetition rate of the experiment). The experiment used to determine T2 is called a 2-Pulse ESEEM. In this experiment the time between two $\pi/2$ pulses is increased. Due to dephasing, the intensity of the echo exponentially decays with increasing time between the pulses since less magnetisation can be refocused (Figure 4.7). This enables determination of the maximum time between the two pulses in the DEER experiment, which will be the longest time between the pulses that still gives a signal.

4.2.4 Domain Topology of the Spt5 NGN and KOW Domains

Previous attempts to crystallise full-length *M. jannaschii* Spt4/5 have not been successful. However, the NGN domain of Spt5 has been crystallised in complex with Spt4 (Hirtreiter et al., 2010a). From the Spt4/5 NGN structure, a homology model of the full-length protein was generated, based on the structure of full length NusG from the thermophilic bacteria *Aquifex aeolicus* (PDB 1M1G). Subsequently, a structure of the full length *Pyrococcus furiosus* Spt4/5

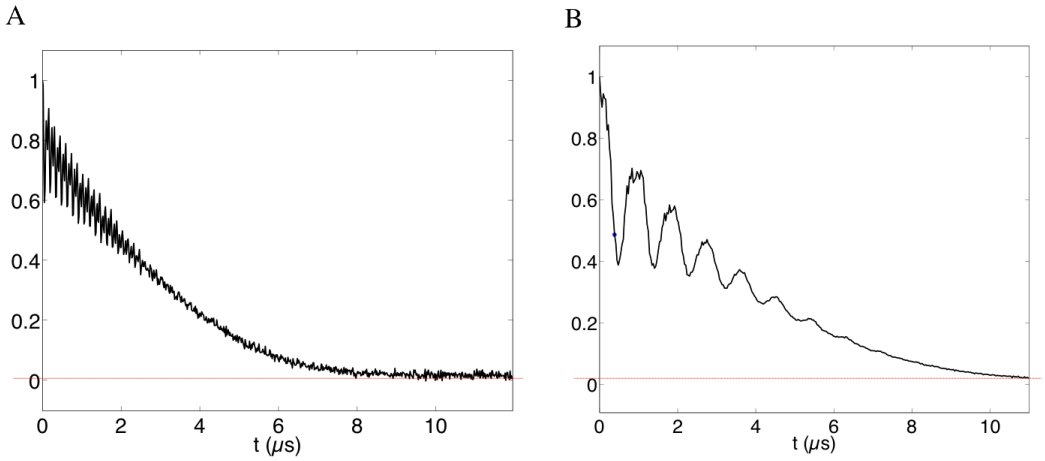


Figure 4.7: Examples of a 2-Pulse ESEEM recorded in A) H_2O-based buffer, and B) 50% D_2O-based buffer. The signal decays inverse exponentially. In H_2O-based buffer there is signal until 8 μs. When there is 50% D_2O in the buffer there is still some signal left at 10 μs.

was obtained (Klein et al., 2011). This crystallised as a dimer. Since Spt4/5 had never been reported to exist in dimeric form in solution it is therefore possible that the domain topology reported in the crystal structure does not represent a solution structure and is artefactual. Furthermore it has become apparent that, like RfaH, NusG from the hyperthermophilic bacteria *Thermotoga maritima* is capable of existing both an inactive, closed conformation and in an open, active conformation (Drögemüller et al., 2013).

In order to determine whether *M. jannaschii* Spt4/5 also exhibits flexibility between the Spt5 NGN and KOW domains, site-directed spin labelling of the NGN and KOW domains followed by DEER was performed. In addition, the distance distribution between the two domains was compared in the presence and absence of Spt4 to determine whether Spt4 influences the domain topology of Spt5.

4.2.5 Predicting the Distance Distributions of Spt5 NGN and KOW Domains

Positions for mutagenesis to generate double-cysteine mutants for site-directed spin labelling were selected to be surface accessible and to give distances that are in the optimal region for

analysis by DEER. Surface accessibility and expected distance distributions were predicted in MMM (version 2013), an open-source program that implements a rotamer library approach to predict the possible conformations of a spin label attached to any position of a protein (Polyhach et al., 2011). Based on these positions, the following double cysteine mutant pairs were selected for analysis: G51/E137, E74/R111, and R111/E137 (Figure 4.8). The positions selected encompass a range of secondary structures. Residues G51 and E74 are within an α helix, and residues R111 and E137 are within a β sheet. These positions are predicted to be relatively rigid with fewer rotamer positions than if they were located on a loop or a turn.

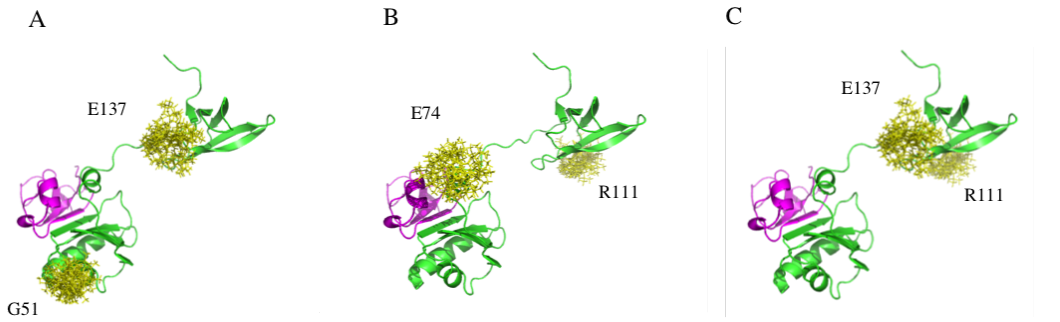


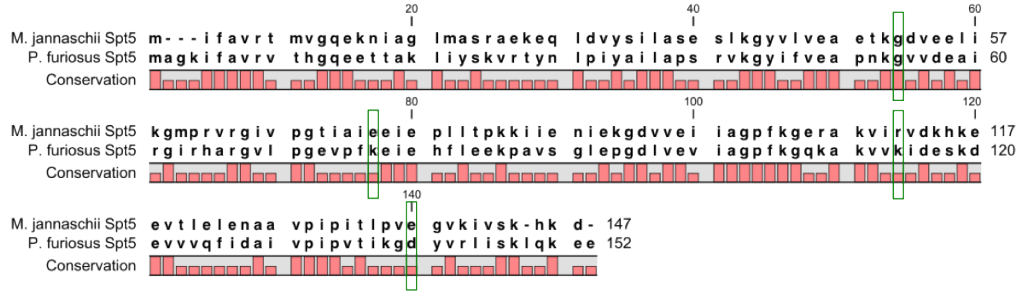
Figure 4.8: Rotamer libraries for *M. jannaschii* Spt4/5 homology model. Double cysteine mutants labelled with iodoacetamide-PROXYL. A) G51/E137, B) E74/R111, C) R111/E137. Green: Spt5. Purple: Spt4. Yellow: Iodoacetamide-PROXYL.

4.2.6 Comparison of the *M. jannaschii* Homology Model and *P. furiosus* Crystal Structure

The *M. jannaschii* and *P. furiosus* Spt5 protein sequences were aligned (Figure 4.9A). The *P. furiosus* Spt4/5 PDB file was loaded into MMM, and the residues corresponding to those selected for DEER analysis in the *M. jannaschii* protein were labelled *in silico* (Figure 4.9B). The expected distance distributions that would be obtained by performing DEER on *P. furiosus* Spt4/5 according to the crystal structure constraints were predicted in MMM. The predicted distributions from *M. jannaschii* and *P. furiosus* were then overlaid, enabling a comparison of the *M. jannaschii* homology model and the *P. furiosus* crystal structure.

Although a crystal structure gives only a single distance between two given atoms, EPR gives

A



B

***M. jannaschii* double mutant**	**Corresponding *P. furiosus* mutant**
G51/E137	G54/D140
E74/R111	K77/K114
R111/E137	K114/D140

Figure 4.9: A) Sequence alignment between *M. jannaschii* (top) and *P. furiosus* (bottom) Spt5. Residues selected for mutation are indicated in a green box. Alignment performed using CLC Main Workbench 5. B) Table showing the three *M. jannaschii* double cysteine mutants and their corresponding mutants in *P. furiosus* Spt5.

a distance distribution. This is calculated based on all of the possible conformations of the attached spin label, which is itself influenced by protein secondary structure and steric factors. Since MMM analysis attaches spin labels to PDB files *in silico*, both the crystal structure and homology model result in predicted distance distributions, and not just a single distance. The two models are in very good agreement, with the plots overlaying significantly. For the G51/E137 mutant (Figure 4.10A) there is significant overlap between the two models with a main population at around 4 nm. The homology model has a slight tendency towards lower distributions, compared to the crystal structure. For the E74/R111 mutant (Figure 4.10B) both the crystal structure and homology mode result in a broad distribution prediction, from 2 nm to 5.5 nm, with a major peak at 4 nm. For the R111/E137 mutant (Figure 4.10C) both the crystal structure and the homology model give a predicted distribution from 1 nm to 3.5 nm.

4.2.7 Labelling of Spt5 Double Cysteine Mutants

Double cysteine mutants of Spt5 were expressed and labelled with iodoacetamide-PROXYL on a NiNTA column. Spt5 was eluted in an imidazole gradient. Unbound spin label was removed and

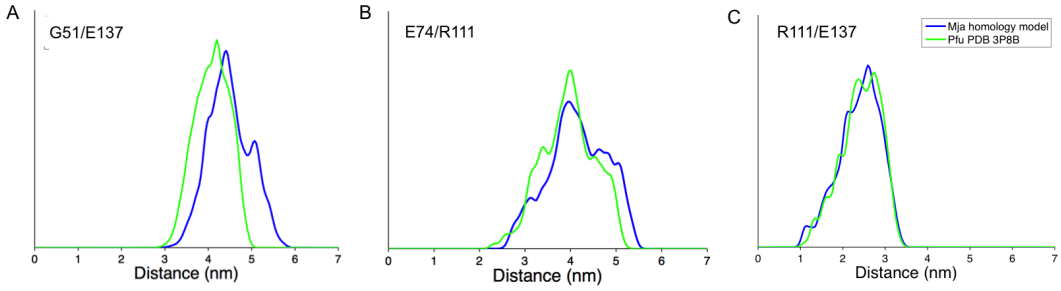


Figure 4.10: Comparison of the *M. jannaschii* Spt4/5 homology model and the *P. furiosus* Spt4/5 crystal structure. The PDB coordinates were loaded into MMM, spin labels attached to the appropriate sites *in silico*, and the distance distributions predicted. A) G51/E137, B) E74/R111, C) R111/E137. Blue: *M. jannaschii* homology model. Green: *P. furiosus* crystal structure. The predicted distance distributions of the two structures overlay well and are in good agreement.

half of the protein was dimerised with Spt4 and repurified on a NiNTA column (Figure 4.11). The protein was concentrated and subsequently diluted two-fold in 50% D8 glycerol in D_2O. CW EPR was conducted on the mutants to determine whether they had been successfully labelled. All of the mutants gave spectra that showed broadening in the high field peak, indicating that they were well-labelled (Figure 4.12).

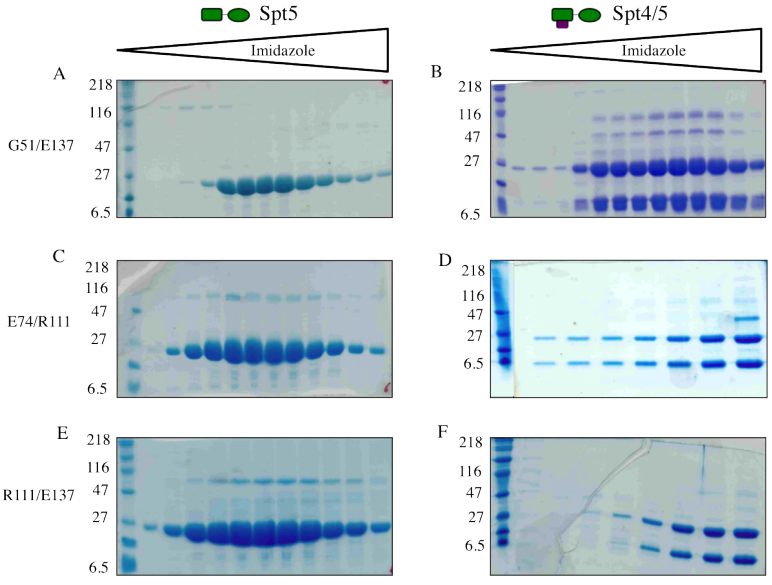


Figure 4.11: SDS PAGE showing purification of His-tagged double cysteine Spt4/5 mutants. A) Spt5 G51C/E137C, B) Spt4/5 G51C/E137C, C) Spt5 E74C/R111C, D) Spt4/5 E74C/R111C, E) Spt5 R111C/E137C, F) Spt4/5 R111C/E137C.

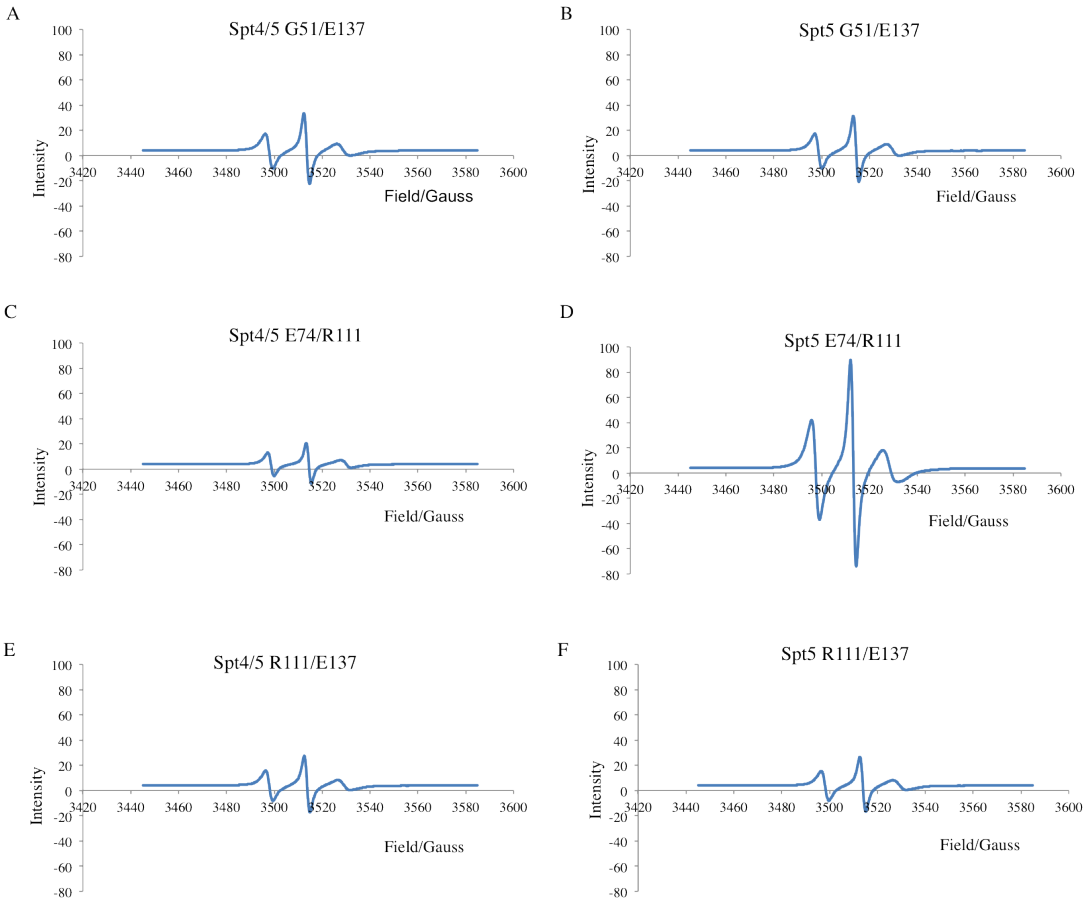


Figure 4.12: CW EPR measurements of double cysteine mutants.

4.2.8 DEER Reveals that Spt4 Reduces Spt5 Flexibility

DEER spectra of double labelled Spt5 and Spt4/5 samples were recorded at 50 K on a *Bruker* ELEXSYS E580 spectrometer at 9 GHz, with an ER-4118-x-MS-3W resonator. The data was analysed in DeerAnalysis (Jeschke et al., 2006) and plotted in MATLAB. For the G51/E137 mutant, Spt4/5 gives one prominent and sharp distribution at 5.8 nm (Figure 4.13A). This is longer than both the major predicted distance at 4 nm and the predicted sub-population at 5.5 nm, based on the homology model. This supports the notion that Spt5 exists in conformations not seen in the crystal structure. When the Spt5 G51/E137 mutant is analysed in the absence of Spt4, two distinct populations are observed (Figure 4.13B). The long distance population at 5.8 nm is still present. In addition there is a significant population that gives a broad distribution

between 3 nm and 5 nm. This second population is in broad agreement with the MMM-predicted distribution, although slightly shorter. For the E74/R111 mutant, there are multiple distances over a broad range for both Spt4/5 and Spt5, indicating that there is backbone flexibility within Spt5 (Figure 4.13C&D). The range of distances observed is approximately 2 nm greater than that predicted by MMM. The distinct populations are better defined in Spt4/5, suggesting that Spt4 restricts mobility. For the R111/E137 mutant there is one major population at 2.6 nm for both Spt4/5 and Spt5 (Figure 4.13E&F). When Spt4 is removed as very small population of 4 nm is present. This suggests that the presence of Spt4 may affect not only the conformation of the NGN domain, but also the KOW domain. Together, the DEER data demonstrates that there is flexibility between the NGN and KOW domains of Spt5. The presence of Spt4 restricts this mobility by stabilising certain conformations.

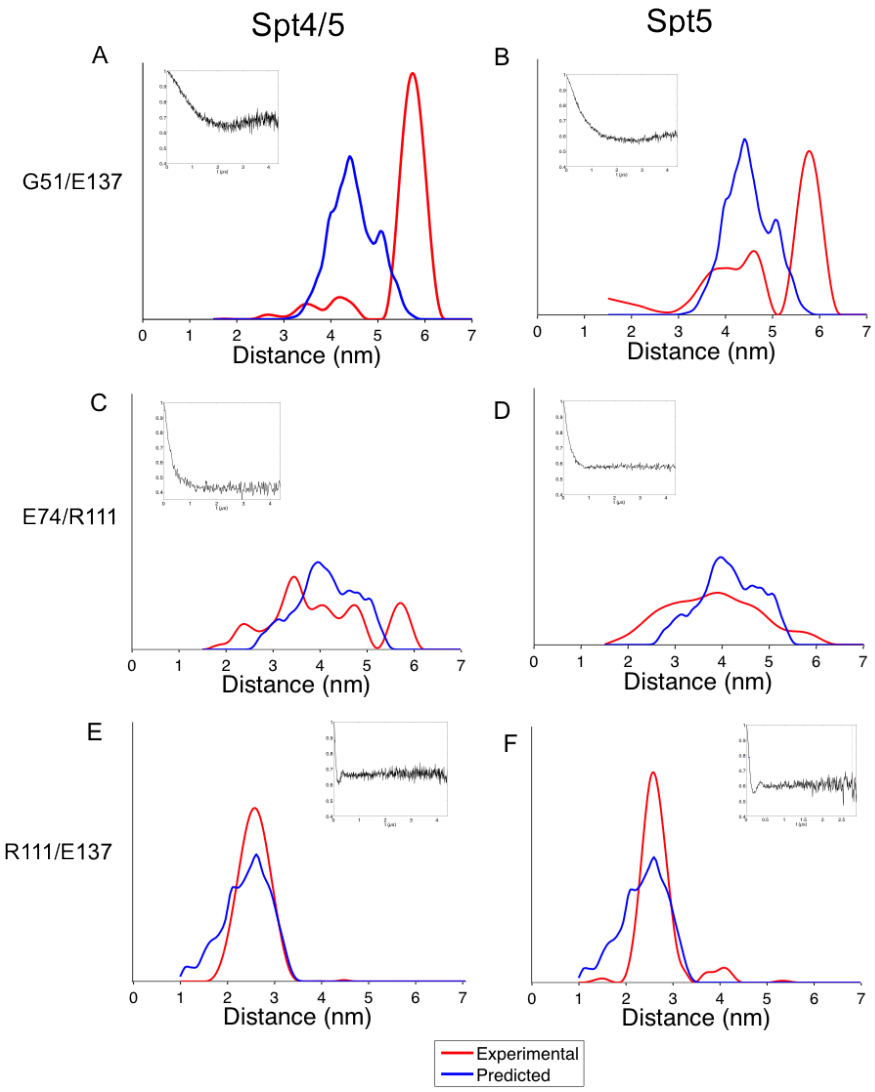


Figure 4.13: DEER analysis of Spt4/5 (left) and Spt5 (right). A&B) G51/E137 mutant. Spt4/5 shows one main distance distribution at 5.8 nm. Spt5 shows two prominent populations of distance distributions, the first between 3 and 5 nm and the second at 5.8 nm. C&D) E74/R111 mutant. Spt4/5 and Spt5 both give a very broad distribution from 2–6 nm. Spt5 shows greater flexibility than Spt4/5, with distinct populations less well defined. E&F) R111/E137 mutant. Spt4/5 and Spt5 both result in a major population at 2.5 nm. When Spt4 is removed a small subpopulation at 4 nm appears. Form factors shown in black. Red: Experimentally obtained data. Blue: Predicted distributions from MMM analysis.

4.3 Domain Dissection of Spt4/5

4.3.1 Spt4 Increases the Thermostability of the Spt5 NGN Domain

In archaea and eukaryotes, the NGN domain of Spt5 is bound by Spt4. Since bacteria have no Spt4 homologue one of the key questions to be addressed is its function. This may also give insights into the evolution of NusG-like proteins. In order to characterise the thermostability of the different domains of Spt4/5, thermofluor assays were carried out. The thermofluor assay is used to ascertain the best buffer conditions (e.g. pH, salt concentration) for stabilising a protein, either for storage or as a precursor to crystallisation trials. However, it also has potential applications in the study of protein-ligand interactions, as a preliminary step in the drug discovery process (Lo et al., 2004).

In the thermofluor assay, a protein is incubated with a fluorescent dye whose fluorescence emission is quenched when within an aqueous buffer but which exhibits high fluorescence when bound to non-polar substances, such as the residues on the interior of a protein. The protein-fluorescent dye mixtures are placed in clear PCR tubes inside a Real Time PCR machine (Niesen et al., 2007). The temperature is increased in 0.5°C increments and the fluorescence emission monitored. Initially, within the aqueous solution where the fluorescence is quenched there is little fluorescence observed. If the protein denatures its hydrophobic core is exposed. The fluorophore binds to the exposed hydrophobic surface resulting in an increase in fluorescence emission (Ericsson et al., 2006). This is seen as a sigmoidal melting curve as the fluorescence emission increases. This may be followed by a decrease as the exposed hydrophobic residues aggregate together, to the exclusion of the dye.

Expression and Purification of Spt4/5 Domain-Deletion Mutants

The Spt4/5, Spt5, Spt4/5 NGN, Spt5 NGN, and Spt5 KOW proteins were expressed as GST fusion proteins and purified on a glutathione column (Figure 4.14). The GST was cleaved from the Spt5 protein overnight and the GST was subsequently removed by heat-inactivation followed by centrifugation (Figure 4.15).

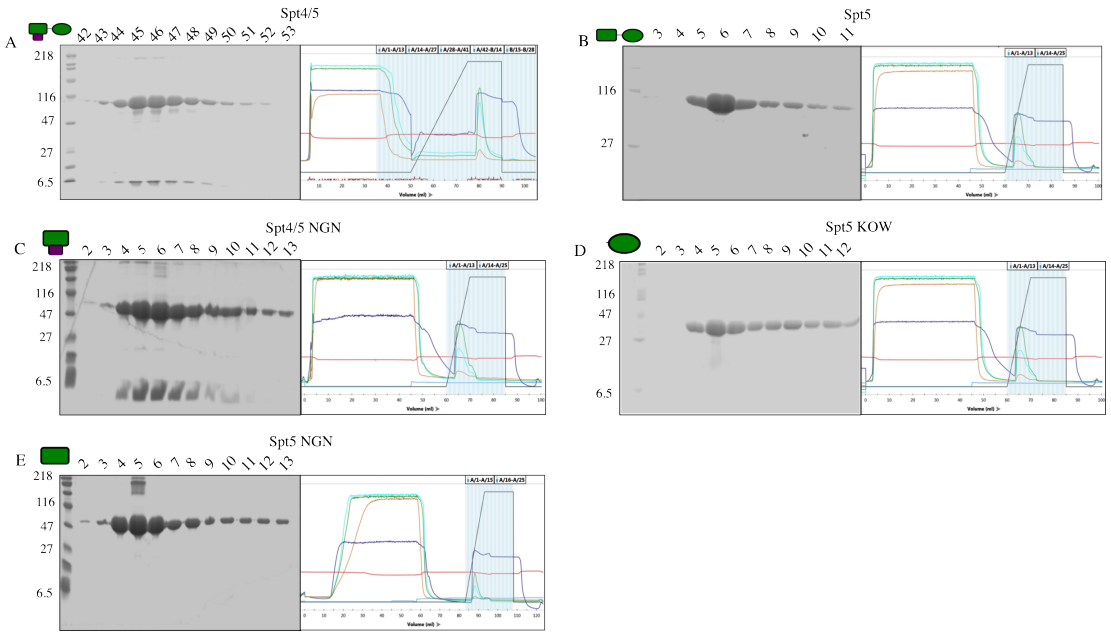


Figure 4.14: Purification of GST-tagged Spt4/5 domain-deletion mutants showing SDS PAGE and chromatograms. A) Spt4/5, B) Spt5, C) Spt4/5 NGN, D) Spt5 KOW, E) Spt5 NGN.

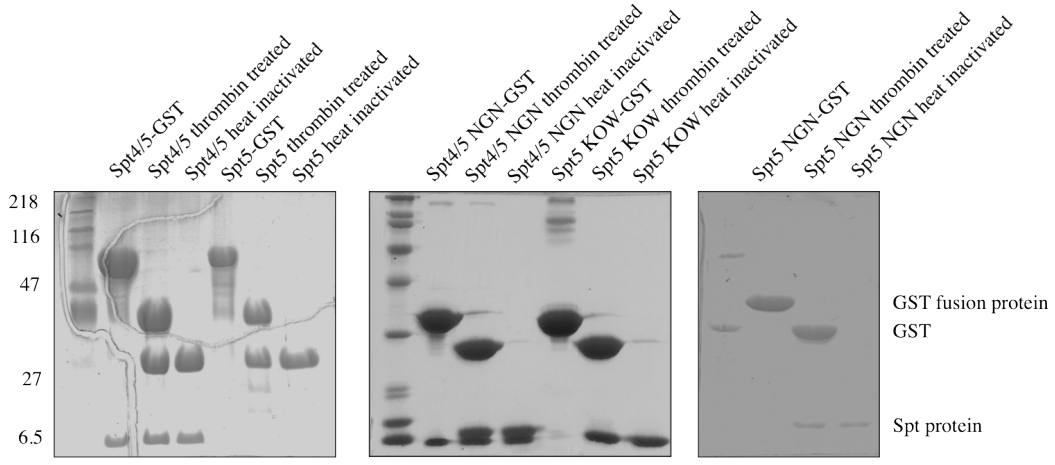


Figure 4.15: SDS PAGE showing thrombin treatment and heat-inactivation of GST-tagged purified Spt4/5 domain deletion mutants. Purified GST-tagged proteins were treated with thrombin overnight to separate the Spt4/5 mutant from GST. The thrombin-treated sample was subsequently heated at 65°C for 30 minutes, resulting in the precipitation of the GST. The sample was then centrifuged to separate the soluble and insoluble fractions, with the Spt4/5 domain deletion mutants being in the soluble fraction.

The domain-deletion variants of Spt4/5 were analysed by thermofluor, and the fluorescence intensity was plotted against the temperature. SYPRO orange was used as the fluorescent dye. This is due to the fact that it has a good signal:noise ratio. When it binds to denatured hen egg lysozyme, it fluoresces with close to 500% the intensity compared to when it is in aqueous buffer only (Niesen et al., 2007). The SYPRO orange was excited at 492 nm and the fluorescence emission was detected at 610 nm. The thermostable Spt4/5 does not denature when heated to 98°C. This is consistent with it being a hyperthermophilic protein (Figure 4.16A). Similarly Spt4/5 NGN (Figure 4.16C) and Spt5 KOW (Figure 4.16E) are also completely thermostable at 98°C. In contrast Spt5 (Figure 4.16B) and the NGN domain of Spt5 (Figure 4.16D) are heat-labile, and melt at 85°C, as indicated by the sigmoidal melting curve. The variants that melt all contain the Spt5 NGN domain but do not contain Spt4. Therefore Spt4 stabilises the NGN domain of Spt5, making it more resistant to thermal denaturation.

Although the melting curves are clear, the increase in fluorescence is lower than the theoretical maximum. When Spt5 melts, the fluorescence increases from approximately 300 RFU to approximately 600 RFU, an increase of 100%. When the Spt5 NGN melts, the fluorescence increases from approximately 500 RFU to just over 1000 RFU, an increase of 100%. This may be because in (Niesen et al., 2007), where the melting of hen egg lysozyme is analysed, 10 mM 3-(cyclohexylamino)-1propanesulfonic acid, 150 mM NaCl was used as the buffer, whereas N500 was used for the analysis of the Spt4/5 domain-deletion mutants. For both Spt5 and Spt5 NGN, the fluorescence emission intensity decreases after reaching its maximum, which is indicative of the aggregation of the denatured protein.

4.3.2 Expression and Purification of Spt4

Spt4 was expressed in Rosetta 2 cells from a pET151D vector with an N-terminal hexa-histidine tag. Spt4 was purified on a NiNTA affinity column and the resulting protein was pure and highly enriched (Figure 4.17).

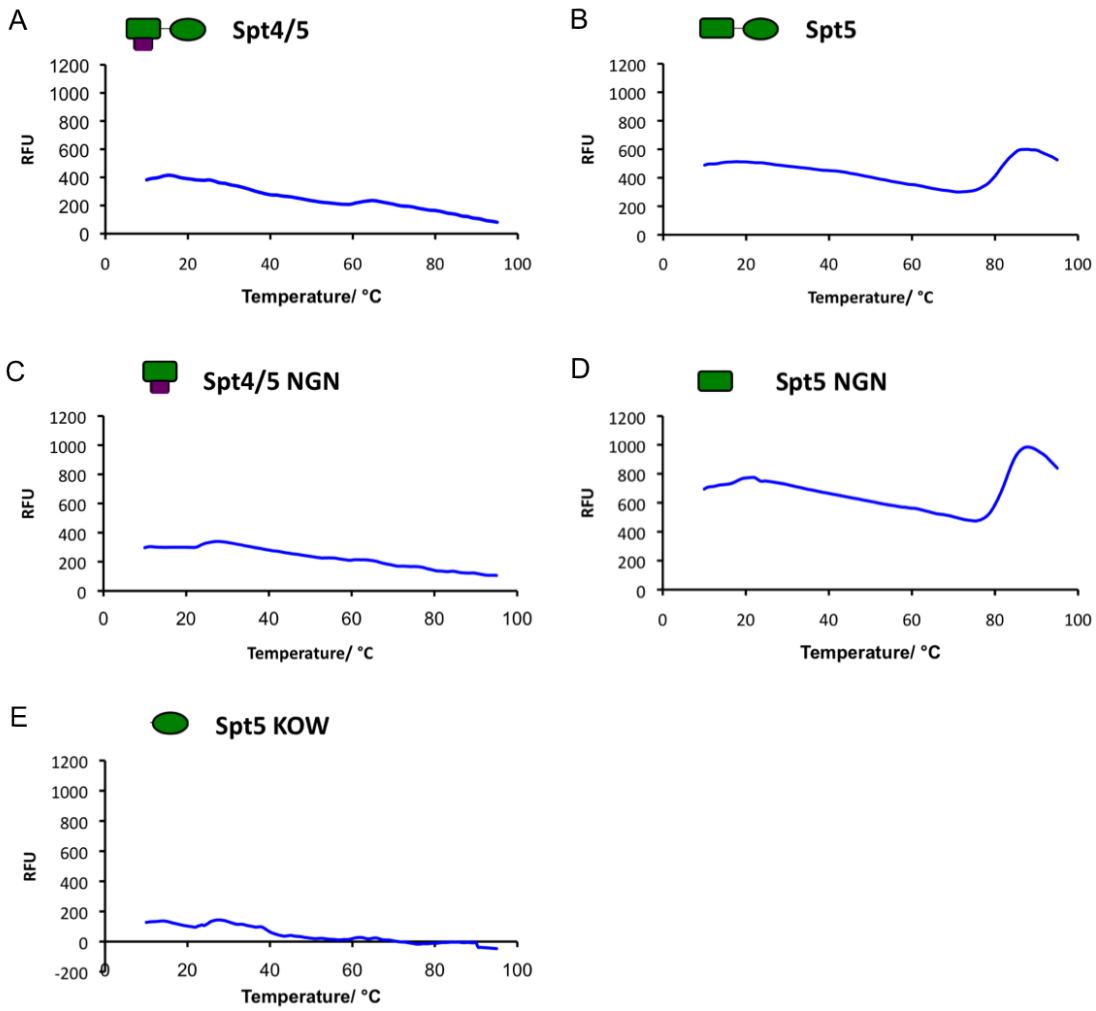


Figure 4.16: Thermofluor analysis of Spt4/5 domain-deletion mutants. A) Spt4/5 is thermostable and does not denature, B) Spt5 denatures at 85°C, C) Spt4/5 NGN is themostable and does not denature, D) Spt5 NGN denatures at 85°C, E) Spt5 KOW is thermostable and does not denature. RFU is the relative fluorescence unit.

4.3.3 Spt4 Protects Spt5 from Proteolysis

The susceptibility of Spt5 and Spt4/5 to proteolytic digestion was analysed by a limited proteolysis approach. In a limited proteolysis assay, protein of interest is incubated with varying concentrations of protease for a defined time. The products of proteolysis are analysed by SDS PAGE (Cleveland et al., 1977). Globular domains of proteins are typically fairly resistant to proteolysis, with proteases preferring disordered or flexible regions as substrates (Fontana et al., 2004). Some studies suggest that a region of at least 12 flexible amino acids is required for

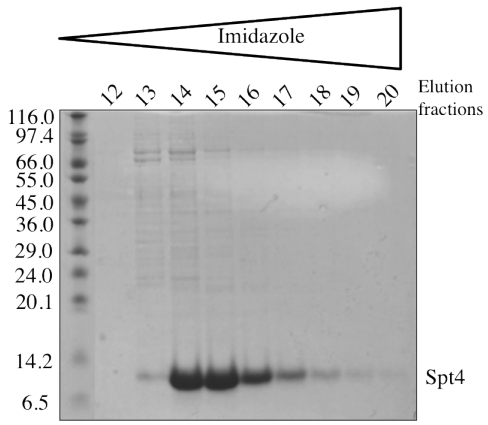


Figure 4.17: SDS PAGE showing the purification of His-tagged Spt4 from a NiNTA column.

proteolysis to occur (Hubbard, 1998). Consequently it is not necessarily expected that protease cleavage is determined by the protease specificity alone, but rather the local protein conformations and flexibility (Fontana et al., 2004). Due to the preference for flexible or disordered regions, limited proteolysis may be used on proteins that are proving difficult to crystallise. This will remove the disordered regions and thereby increase the likelihood of forming crystals.

In order to analyse the function of Spt4, limited proteolysis was performed on Spt4/5 and Spt5 using trypsin as the protease. Spt4/5 is resistant to proteolysis by trypsin even at the highest protease concentrations (Figure 4.18A). Spt5 alone, in contrast, is cleaved at 12.5, 25, and 50 μg/ml trypsin into two fragments (Figure 4.18B). This indicates that Spt5 has greater structural integrity when bound to Spt4. Analysis of the *P. furiosus* Spt4/5 crystal structure provides insight into how Spt4 may protect Spt5 from proteolysis. In addition to binding the NGN domain of Spt5, Spt4 also occludes the flexible linker between the NGN and KOW domains (Figure 4.18C). It is therefore likely that Spt4 protects the flexible linker of Spt5. Trypsin cleaves proteins after arginine and lysine residues. The NGN domain of Spt5 contains four arginines and six lysines. The KOW domain of Spt5 contains two arginines and seven lysines. The linker between the two domains contains one lysine (K84) (Figure 4.19). It is clear that not every positive residue of Spt4/5 is susceptible to cleavage, as this would result in many small protein fragments which would be too small to visualise on a 16% acrylamide gel. Since DEER has shown that there is flexibility between the two domains, and proteases have a tendency to cleave flexible linker regions of proteins, it is likely that the lysine in the linker is the positive residue most susceptible

to cleavage and the substrate of trypsin. This could be investigated through N-terminal Edman sequencing, or by mass spectrometry of the digested Spt5 fragments. Alternatively K84 could be mutated to a non-polar residue and the assay repeated to determine whether proteolysis still occurs in the same manner.

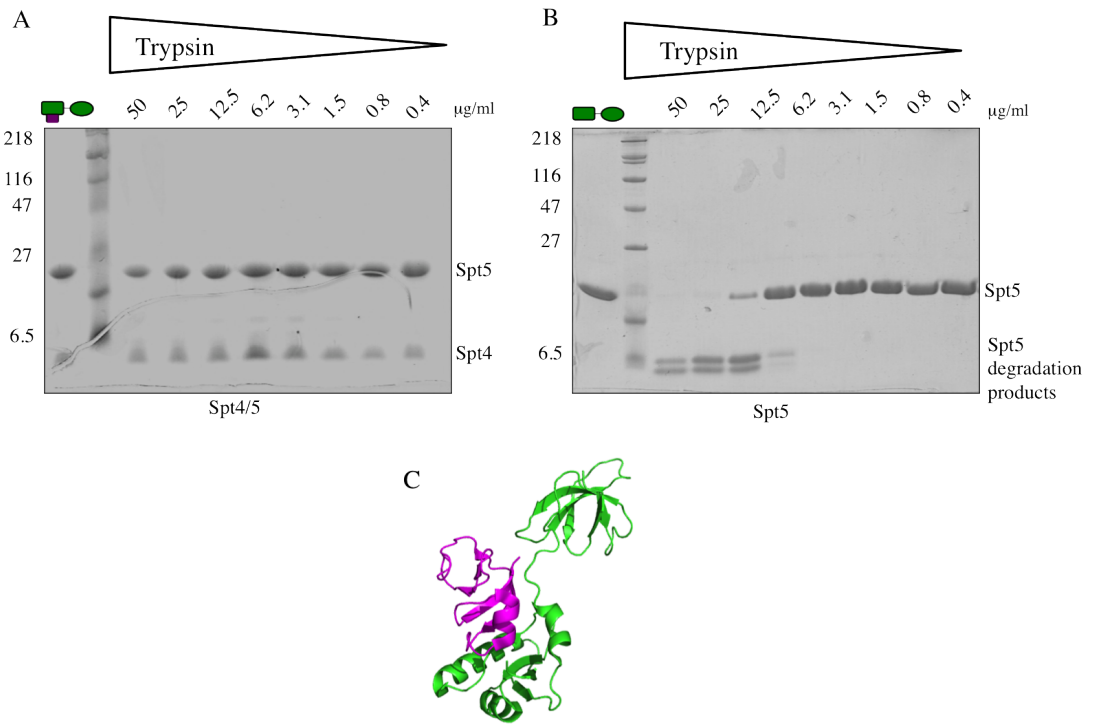
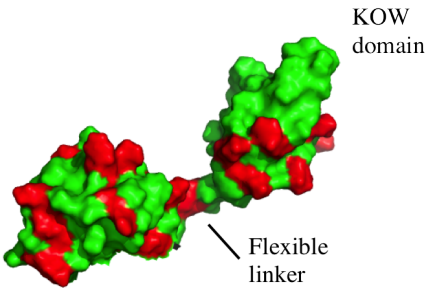


Figure 4.18: Limited proteolysis of A) Spt4/5 and B) Spt5. Trypsin concentrations in μg/ml indicated above lanes. Spt4/5 is resistant to trypsin cleavage, even at the highest concentrations of protease. In contrast, Spt5 alone is cleaved into two fragments at the three highest trypsin concentrations. C) Spt4 may protect the flexible linker of Spt5 from proteolysis.

4.3.4 Spt4 is Stabilised by a Cys4 Zinc Finger

Spt4 contains four conserved cysteines that coordinate a zinc ion (Hirtreiter et al., 2010b) (Figure 4.20). Limited proteolysis of Spt4 was conducted in the absence and presence of EDTA to explore the role of Zn^{2+} coordination on its stability. Addition of EDTA at a final concentration of 250 mM causes chelation of the Zn^{2+} ions. There are eight lysines and one arginine in Spt4 that are possible substrates for trypsin. Spt4 is completely resistant to degradation by trypsin (Figure 4.21A). When EDTA is added, Spt4 becomes susceptible to proteolysis (Figure 4.21B).

A



B

1	MIFAVRTMVG	QEKNIAGLMA	SRAEKEQLDV	YSILASESLK	40
41	GYVLVEAETK	GDVEELIKGM	PRVRGIVPGT	IAIEEIEPLL	80
81	TPKKIIENIE	KGDVVEIIAG	PFKGERAKVI	RVDKHKEEVT	120
121	LELENAAVPI	PITLPVEGVK	IVSKHKD	147	

Figure 4.19: A) Space-filling *M. jannaschii* Spt5 homology model with positively charged lysine and arginine residues coloured in red. B) Amino acid sequence of *M. jannaschii* Spt5 with positive residues coloured in red. K84, which is positioned within the flexible linker, is probably the target for trypsin.

Since proteases typically degrade flexible regions of proteins, it stands that when the Zn^{2+} is removed, Spt4 becomes more flexible or loses structural integrity. Therefore the Zn^{2+} is required to maintain the structural integrity of Spt4.

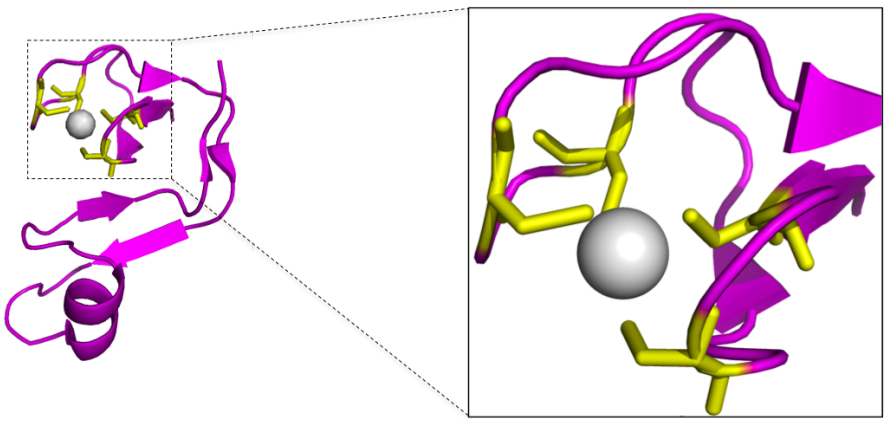


Figure 4.20: Coordination of zinc by Spt4. Four conserved cysteine residues coordinate a zinc atom. *M. jannaschii* Spt4 shown from PDB file 3LPE. Purple: Spt4. Yellow: Cysteines of Spt4. Grey: Zn.

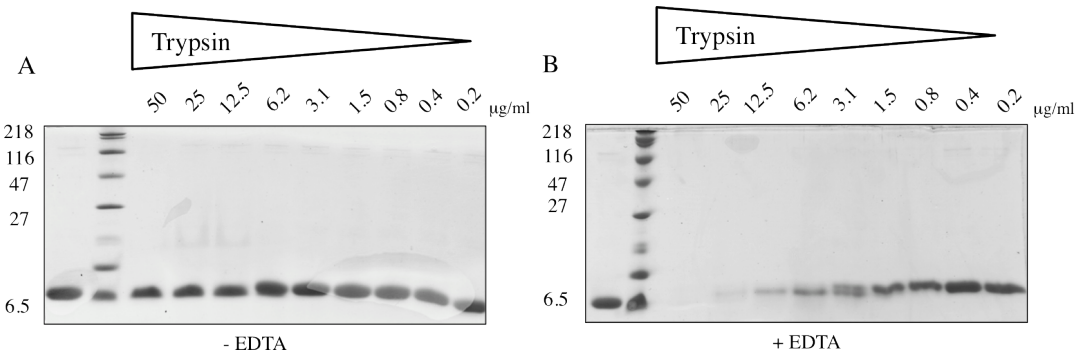


Figure 4.21: Limited proteolysis of Spt4 in the A) absence, and B) presence of EDTA. Addition of EDTA results in the chelation of Zn^{2+}, and Spt4 is degraded. When the Zn^{2+} remains coordinated by the cysteines, Spt4 is resistant to trypsin degradation at the highest concentrations of protease.

4.3.5 Discussion

In this chapter a biophysical characterisation of Spt4/5 has been performed. An SDSL approach followed by DEER was carried out whereby the Spt5 NGN and KOW domains were each labelled. This revealed that Spt5 is able to exist in at least two conformations in solution, as evidenced by distinct distance distributions. Spt4 influences the relative proportion of the populations. This raises the possibility that the two conformations have different functions and that Spt4 may regulate their interconversion, for example by switching Spt5 from a RNAP processivity-enhancing factor to transcription-translation coupling factor. The Spt4-free Spt5 would possibly rely upon interactions with additional proteins to stabilise the NGN domain. Alternatively one of the conformations of Spt5 may be inactive and the other active in a similar manner to RfaH.

Distances between two points on a protein can also be determined by the incorporation of fluorescent labels into the protein, where the emission spectrum of one label overlaps with the absorption spectrum of the second label. Excitement of the first label will result in resonance transfer of energy to the second label, with the efficiency of transfer being inversely proportional to the distance. This experiment is called FRET (fluorescence resonance energy transfer). FRET and EPR typically give results in good agreement with one another (Grohmann et al., 2010; Klose et al., 2012). One advantage of SDSL, however, is the small size of spin labels. 3-(2-Iodoacetamido)-PROXYL is not much bigger than an amino acid (Figure 4.3), whereas fluorescent labels are often large, bulky, and aromatic. Therefore incorporation of a nitroxide

spin label should result in minimal perturbation of protein structure or function.

One possibility for further studying the role of Spt4 would be to analyse the relative levels of Spt4 and Spt5 gene expression *in vivo* to determine whether they are expressed with equal stoichiometry, or whether there is an excess of Spt5. This could be determined by quantitative Western blotting using antibodies against Spt4 and Spt5. It would also be interesting to compare expression levels from mid log and stationary-phase cells to determine whether Spt4 and Spt5 expression levels are regulated differently. Additionally, knocking out Spt4 and analysing the effects on global gene expression would allow determination of whether Spt5 is able to efficiently stimulate gene expression in the absence of Spt4 and whether Spt4 regulates the expression of distinct subsets of genes.

A thermofluor assay was performed on domain-deletion variants of Spt4/5. This indicated that the NGN domain of Spt5 is stabilised when bound to Spt4. This function of Spt4 is consistent with *M. jannaschii* being a thermophile that grows at elevated temperatures. However, there are bacteria which also grow at extreme temperatures that do not have Spt4. In addition eukaryotic Spt5 is also bound by Spt4, although most eukaryotes do not grow at high temperatures. This suggests that Spt4 may have additional functions. It is positioned close to where the DNA strands anneal within the elongation complex and it may therefore have a role in this process or in contacting the upstream DNA (Werner, 2013).

Limited proteolysis was used to study the susceptibility of Spt5 to tryptic digestion. Spt4/5 is resistant to digestion. However Spt5 alone is cleaved into two fragments. Analysis of the *P. furiosus* Spt4/5 crystal structure leads to the conclusion that Spt4 likely protects the Spt5 flexible linker. This agrees with the DEER data in suggesting that Spt4 reduces the conformational flexibility between the Spt5 NGN and KOW domains.

Chapter 5

Analysis of the RNA-binding Properties of NusA

5.1 Introduction

5.1.1 Structure and Function of NusA

In *E. coli*, *nusA* is an essential gene and homologues have been identified in all sequenced bacterial genomes. Bacterial NusA comprises an RNAP-binding N-terminal domain, an S1 domain, and two KH domains. A crystal structure of *M. tuberculosis* NusA without its N-terminal domain and in complex with an 11-mer RNA has been solved (Beuth et al., 2005). Within this structure the two KH domains bind RNA, but the S1 domain does not. The two KH domains are separated by 30 Å and form a single, continuous RNA-binding surface (Figure 5.1). This is illustrated by the fact that one of the nucleotides within the RNA is contacted by residues from both the KH1 and the KH2 domain. The bound RNA is in an elongated conformation and its position is guided by regions of positive electrostatic charge. The KH domains undergo multiple types of interaction with the RNA, providing sequence specificity. These include hydrogen bonds from both side chains and the backbone, electrostatic interactions, and hydrophobic interactions between non-aromatic side chains and the aromatic RNA bases. Furthermore, *M. tuberculosis* NusA also makes interactions with 2′-OH groups of the ribose rings, thereby enabling it to distinguish between single-stranded DNA and single-stranded RNA. The crystal structure of free NusA has also been solved, revealing that NusA does not undergo a conformational change upon RNA binding (Beuth et al., 2005).

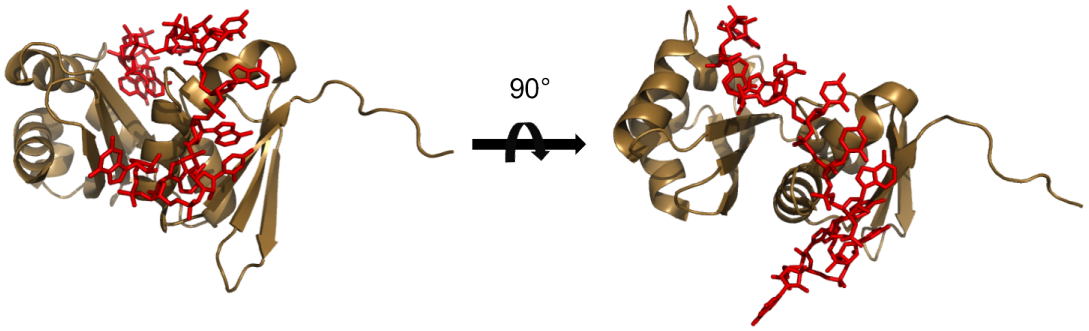


Figure 5.1: *M. tuberculosis* NusA bound to RNA (PDB 2ASB).

5.1.2 Functions of Bacterial NusA

In bacteria, NusA is able to bind to RNAP and promote termination (Schmidt and Chamberlin, 1987). However, when acting in the context of an antitermination complex, NusA has the opposite function, namely enabling RNAP to enter a pause and termination-resistant state. In *E. coli* the host proteins NusA, NusB, NusE (S10), and NusG (Mason and Greenblatt, 1991) assemble towards the 5′ end of the transcript at conserved sequences called BoxA and BoxB, with NusA binding a spacer region between them. Antitermination was first described in bacteriophage λ-infected cells. Phage λ hijacks the host RNAP and Nus factors to enable expression of its own genes. Antitermination also occurs on ribosomal operons, which are not protein-coding. This prevents termination by rho, and enables robust expression of the rRNA.

NusA also functions in the processing of the 16S ribosomal RNA. The ribosome assembles co-transcriptionally, and mutations within NusA prevent efficient ribosome biogenesis and result in a buildup of 21S precursor (Bubunenko et al., 2013). NusA facilitates rRNA folding through constraining it at its 5′ end. NusA binds to RNAP and to a BoxA-like region towards the 5′ end of the transcript. The 3′ end of the rRNA is constrained by the fact that it is within the RNAP active site. Consequently the RNA forms a loop as expression of the rRNA progresses, facilitating its folding.

5.1.3 Whole Genome Occupancy of *M. jannaschii* NusA

ChIP (Chromatin Immunoprecipitation) data obtained by Dr Katherine Smollett in our lab supports a role for archaeal NusA in transcriptional regulation. In ChIP-seq, cells are grown and the protein that is bound to DNA is crosslinked by formaldehyde treatment. The cells are then lysed and the DNA is sheared. The protein of interest is then pulled out by immunoprecipitation and the crosslinks reversed by heat treatment. The DNA that was bound by the protein is sequenced, and the obtained sequences can then be identified within the genome (Johnson et al., 2007).

The data demonstrates that the Spt4/5 occupancy along protein-encoding and non-protein-encoding genes mirrors that of RNAP (Figure 5.2) indicating that Spt4/5 is a general transcrip-

tion factor. NusA is also present across the entirety of the genes. However, its profile does not mirror RNAP and Spt4/5, and it associates with transcription units in a different manner. This could be via NusA-RNAP or NusA-RNA interactions. Alternatively, NusA may be associated indirectly via other proteins. However, the fact that NusA is present across transcription units provides a sound rationale for studying archaeal NusA at a transcriptional level.

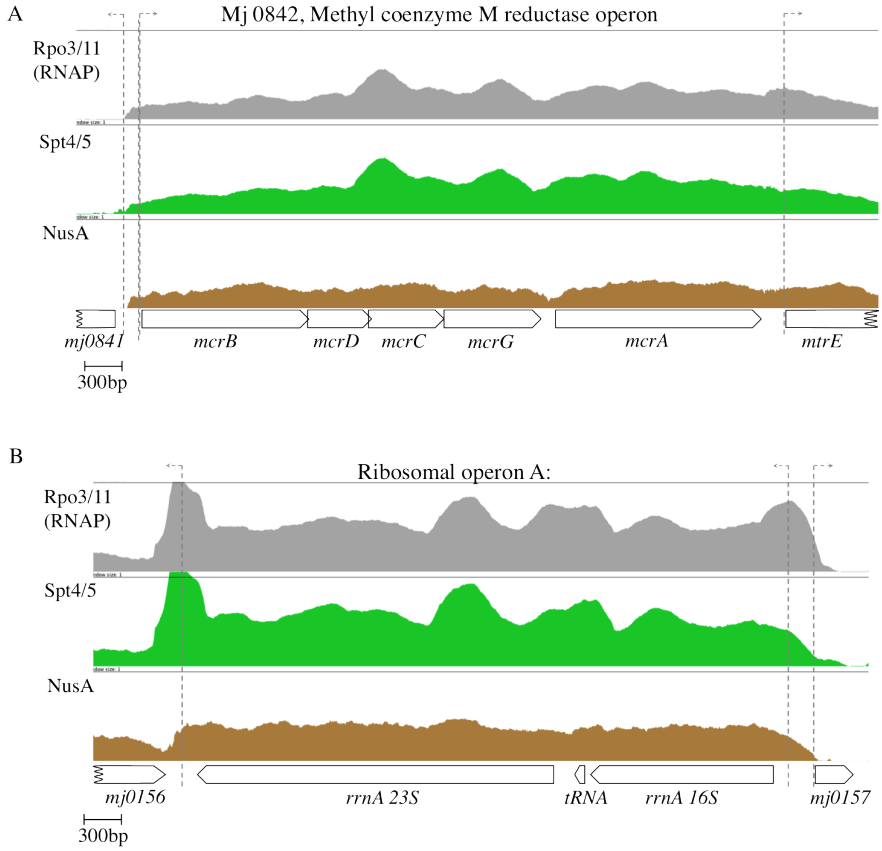


Figure 5.2: Gene occupancy of putative antitermination factors as determined by ChIP. A) The protein-encoding methyl coenzyme M reductase operon. Methyl coenzyme M reductase catalyses the final step in the methane-production pathway. B) The ribosomal operon A, which encodes the RNA component of the ribosome, and is not translated. For both protein-encoding and non-protein-encoding genes RNAP (grey) is present across the entire gene. Spt4/5 (green) closely follows the path of RNAP indicating that it is associated with the elongation complex. NusA (brown) is present along the entirety of protein-encoding and non-protein-encoding genes, but does not follow the RNAP profile. Graphs show Log_2(IP/input) averaged over replicates. RNAP: n=3, Spt4/5: n=3, NusA: n=2.

5.1.4 Chapter Aims

The aims of this chapter are to analyse the RNA binding of *M. jannaschii* NusA. To this end a homology model of *M. jannaschii* NusA was generated. This was to enable prediction as to whether it is likely to adopt a similar fold to previously solved bacterial and archaeal NusA structures. This chapter will analyse the RNA-binding activities of NusA by employing a site-directed mutagenesis approach to determine regions of the protein that are important for RNA binding. A comparison between non-biological A2 RNA and biologically relevant rRNA leader/downstream regions will be made. The cloning of the leader and downstream regions of the *rrnA* gene may also give hints as to whether archaeal NusA is involved in rRNA processing. This chapter also aims to determine whether *M. jannaschii* NusA binds to RNA with any sequence bias, or whether it binds with little discrimination by using a technique called Scaffold-Independent Analysis. This could potentially yield insights into its function. For example, if it were to show preference for a BoxA-like sequence this would suggest a possible antiterminator role.

5.2 Generation of *M. jannaschii* NusA Homology Model

Whilst bacterial and archaeal crystal structures of NusA have been solved, there is no structure available for *M. jannaschii* NusA. The structure of *Aeropyrum pernix* NusA in isolation, and in complex with RNA has been solved (PDB 2CXC) (Shibata et al., 2007). *A. pernix* is a hyperthermophilic crenarchaea that grows at 70–100°C. To gain insight into the possible structure of *M. jannaschii* NusA, the sequences of *M. jannaschii* NusA and *A. pernix* NusA were first aligned. The alignment indicates that the two proteins are homologous and that they share 30% sequence identity. From this it was concluded that *A. pernix* NusA was a good template for generating a *M. jannaschii* NusA model (Figure 5.3A). The homology model was created using SWISS-MODEL[1] (Schwede et al., 2003). The homology model indicates that *M. jannaschii* NusA also comprises two KH domains, each containing two α helices and three β strands. The RNA-binding GXXG motifs are located on the loop between the two α helices in each domain. The model overlays very well with the parent structure (Figure 5.3D).

[1] http://swissmodel.expasy.org/interactive

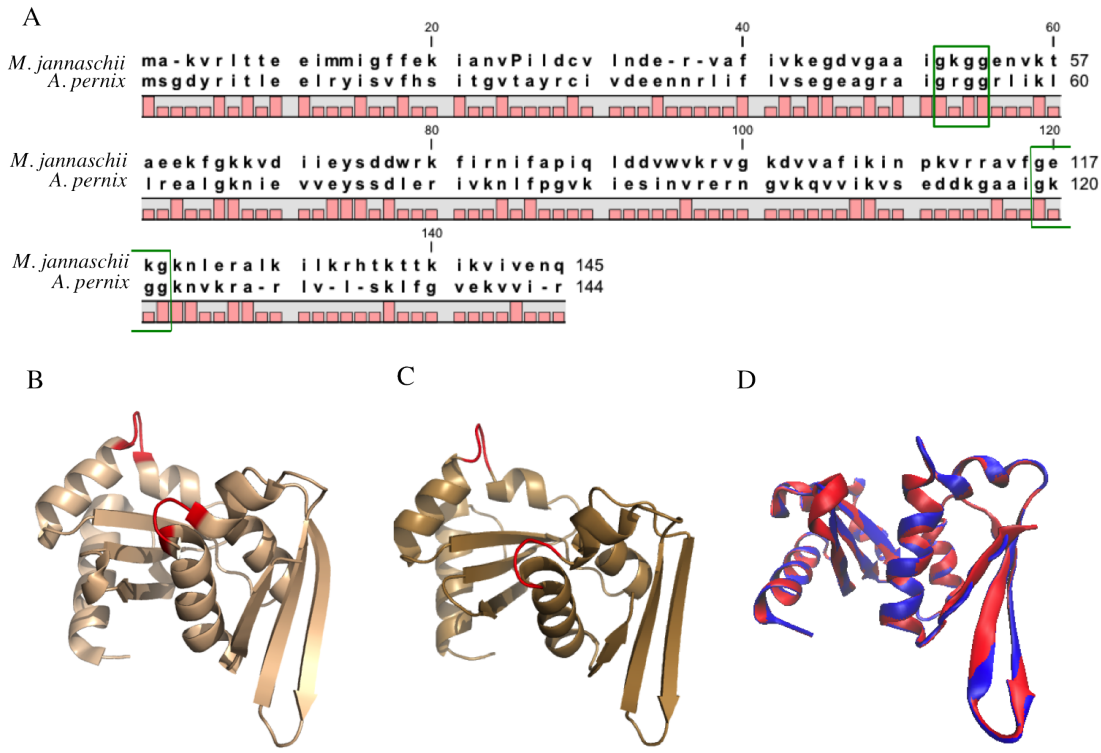


Figure 5.3: Generation of *M. jannaschii* NusA homology model using SWISS-MODEL. A) Sequence alignment between *M. jannaschii* and *A. pernix* NusA. RNA-binding GXXG loops indicated in green boxes. B) Crystal structure of *A. pernix* NusA (PDB 2CXC) used as the parent structure. GXXG loops indicated in red. C) Homology model. D) Overlay of *A. pernix* parent structure (red) and the *M. jannaschii* homology model (blue).

5.3 Analysis of NusA RNA-Binding Mutants

5.3.1 Expression and Purification of NusA Mutants

NusA mutants were expressed in Rosetta 2 cells. Protein was extracted from the cells and the supernatant was heat-inactivated to denature most of the endogenous *E. coli* proteins. The remaining soluble protein was purified on a heparin column (Figure 5.4).

5.3.2 EMSAs Using Fluorescently Labelled A2 RNA

In order to characterise the RNA-binding activity of NusA, electrophoretic mobility shift assays (EMSA) were performed. In an EMSA two putative interacting partners are incubated

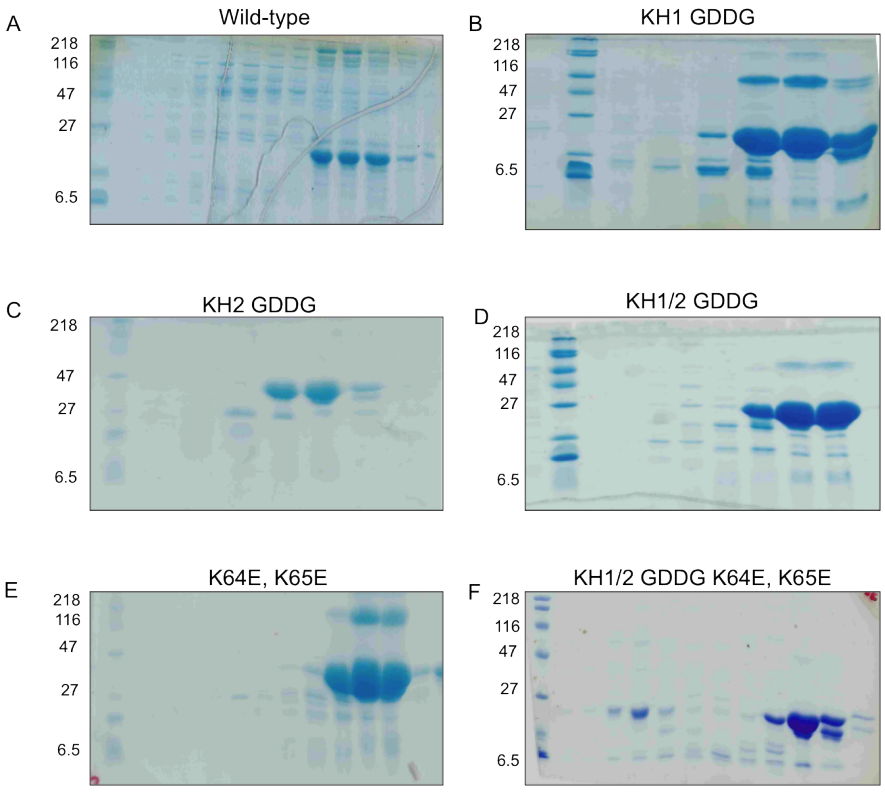


Figure 5.4: SDS PAGE of NusA mutant purifications by heparin affinity chromatography. A) Wild type, B) KH1 GDDG, C) KH2 GDDG, D) KH1/2 GDDG, E) K64E, K65E, F) KH1/2 GDDG K64E, K65E.

together, with the first partner maintained at constant concentration and the second partner titrated in. The incubations are then loaded on to a native polyacrylamide gel and separated by electrophoresis. As no SDS is present the complexes maintain their integrity as they migrate through the gel. Higher molecular weight complexes have lower mobility through the gel than single species, and therefore complex formation is identified by a shift of a band in the gel. Usually one of the partners is radio-labelled or fluorescently labelled to enable detection.

Varying concentrations of NusA mutants were incubated with a constant concentration of Cy3-labelled A2 RNA. In all EMSAs using A2 RNA the RNA concentration was 50 nM, and the protein concentrations were 1 μM, 2.5 μM, 5 μM, 7.5 μM, 10 μM, 15 μM, and 20 μM. When labelled RNA is incubated with wild-type NusA, complex formation is observed at 5 μM demonstrating that NusA is able to bind to RNA (Figure 5.5).

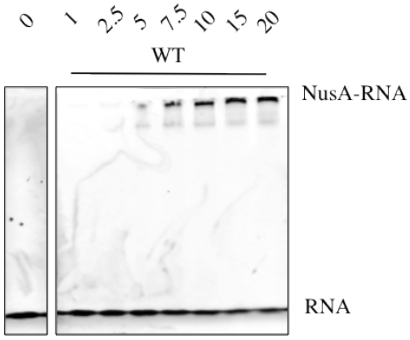


Figure 5.5: NusA bandshifts with A2 RNA probe. The RNA concentration is 50 nM. NusA concentration in μM indicated above individual lanes. Complex formation is observed weakly at 5 μM, and more strongly at 7.5 μM.

Mutational analysis was conducted on NusA to determine the relative contribution of the two KH domains to RNA binding. The GXXG loops of the two KH domains were mutated to GDDG using a splice-by-overlap methodology (see Chapter 2). This is a well-characterised mutation for abolishing the RNA-binding of KH domains without compromising their structural integrity (Hollingworth et al., 2012). Mutation of the KH1 GXXG motif to GDDG (K50, G51–to–D50, D51) impairs the RNA-binding activity of NusA, with complex formation not being observed until the concentration of NusA reaches 15 μM (Figure 5.6 left panel). This indicates that the KH1 domain of NusA contributes to RNA-binding. Mutation of the KH2 GXXG motif to GDDG (E117, K118–to–D117, D118), in contrast, has no effect on the RNA-binding activity of NusA (Figure 5.6 middle panel). Complex formation is observed at 5 μM NusA, which is the same as for the wild-type protein. This demonstrates that the KH2 domain of NusA has little or no role in RNA binding. This is possibly because one of the residues in the KH2 GXXG motif (E117) is already negatively charged. The relative contributions of the two domains is supported by analysis of a double KH mutant where both KH domain GXXG loops have been mutated to GDDG. With the double KH domain mutant, complex formation is observed at 15 μM (Figure 5.6 right panel). This is the same as for the single KH1 mutant.

The double KH mutant was expected to completely abolish RNA binding, based on previous reports of KH domain proteins (Hollingworth et al., 2012). The fact that it still binds RNA suggests that there are other regions of NusA that are involved in the binding of RNA. Further

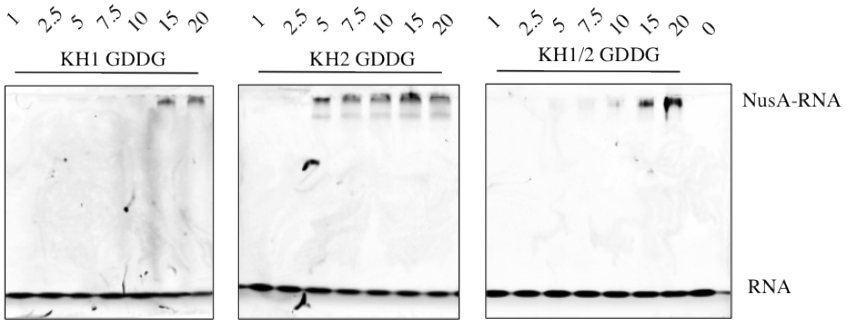


Figure 5.6: NusA bandshifts with A2 RNA probe. The RNA concentration is 50 nM. NusA concentration in μM indicated above individual lanes. Mutation of the KH1 GXXG loop to GDDG results in a strong phenotype. Conversely mutation of the KH2 GXXG loop to GDDG has no effect on RNA binding. Combination of the KH1 and KH2 GXXG loop mutations has the same phenotype as the single KH1 mutant, suggesting that the KH2 domain has no role in RNA binding.

analysis of the NusA homology model identified the presence of two positively charged lysine residues (K64, K65) in the cleft between the two KH domains (Figure 5.7).

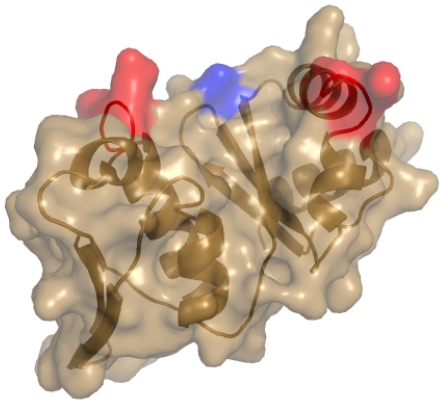


Figure 5.7: Space filling *M. jannaschii* NusA homology model showing the GXXG loops (red) and the two lysine residues K64, and K65 (blue).

It seemed likely that K64 and K65 bind to RNA via electrostatic interactions with the negatively charged ribose-phosphate backbone of the RNA. To test this hypothesis, the two lysines were mutated to glutamate in order to reverse their charge. The mutant was soluble and stable at 65°C. The charge-reversal mutation resulted in a minor phenotype, with slightly impaired RNA

binding (Figure 5.8 left panel), indicating that K64 and K65 are involved in the interaction with RNA. Combination of the double KH domain mutation with the charge reversal mutation completely abolished RNA binding, with no complex formation observed even at the highest concentrations of NusA (Figure 5.8 right panel).

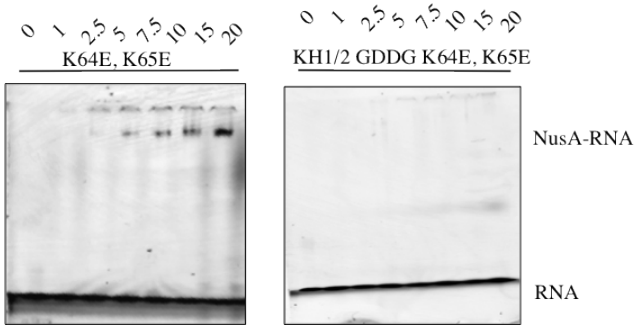


Figure 5.8: NusA bandshifts with A2 RNA probe. The RNA concentration is 50 nM. NusA concentration in μM indicated above individual lanes. Mutation of the K64 and K65 to glutamate mildly impairs RNA binding indicating a role for these two residues in the binding of RNA. Combining the two KH GXXG loop mutations with the double K–to–E mutation completely abolishes the RNA binding activity of NusA.

5.4 Cloning of the *M. jannaschii* 16S rRNA Leader and Downstream Regions

For protein-encoding genes in bacteria, the binding of the KOW domain of NusG to the S10 protein of the ribosome prevents termination by the rho helicase, which competes for binding to the KOW domain. rRNA operons encode the RNA that is incorporated into the ribosome and therefore they are not translated. This means that they are more prone to termination by rho. Antitermination complex formation reduces rho termination across the rRNA operon. Complex formation occurs when NusA binds towards the 5′ end of rRNA as it is being transcribed. NusA is just one component of antitermination complexes, which also include NusG, NusB, and the ribosomal protein S10.

NusA also facilitates the co-transcriptional folding of rRNA. The 3′ end of the transcript is bound within the active centre of RNAP. The 5′ end of the transcript is bound by RNAP-associated

NusA. Therefore, as the RNA chain increases in length, it forms a growing loop which facilitates its correct folding. The 16S rRNA leader and the region downstream of the 16S rRNA gene are complementary in sequence and hybridise to form a region of dsRNA that is cleaved by RNaseIII to liberate the 16S rRNA. In NusA mutants, rRNA processing functions abnormally. It is not known whether archaeal NusA also binds to the *rrn* operon leader regions.

M. jannaschii contains two rRNA operons, called *rrnA* and *rrnB*. The 16S rRNA(A) leader and downstream elements were cloned into a pGEM-T vector. The *rrnA* operon is on the negative DNA strand in the *M. jannaschii* genome. The leader region of the 16S rRNA was cloned starting at position 159,635 and terminating at 159,460 of the *M. jannaschii* genome. The PCR reaction resulted in two products of about 275 bp and 180 bp. This is probably because the two ribosomal leader sequences in *M. jannaschii* have highly similar sequences. The expected size was 175 bp, so the lower mass product was correct. The downstream region of the 16S rRNA gene was cloned, starting at position 157,990 and terminating at 157,910 of the *M. jannaschii* genome. The PCR reaction resulted in one product of 80 bp, as expected (Figure 5.9). In addition, a leader-downstream fusion construct was also cloned using a splice-by-overlap extension methodology (see Chapter 2). This resulted in a product of 255 bp. *Taq* DNA polymerase was used for the PCR reactions as it leaves A overhangs, required for ligation into the pGEM-T vector. Radiolabelled RNA was subsequently prepared by *in vitro* transcription and purified using G25 spin column.

The predicted fold of the cloned rRNA regions was determined using RNAfold (Gruber et al., 2008). RNAfold predicts the secondary structure of RNA and ssDNA based on minimum free energy constraints. Although secondary structure is predicted for the 175 nucleotide leader region, the prediction is not high confidence, indicated by the green colour. This represents about a 50% confidence (Figure 5.10A). The downstream region is only 80 nucleotides long and is predicted to form two short regions of double-stranded RNA, each of about 5 basepairs. The prediction is high confidence, indicated by the red colour (Figure 5.10C). The fusion of the leader and downstream RNA is 255 nucleotides long. An extended region of about 38 basepairs is predicted with very high confidence. The leader and downstream regions are complementary in sequence, suggesting that they undergo Watson-Crick base-pairing during transcription of the 16S rRNA

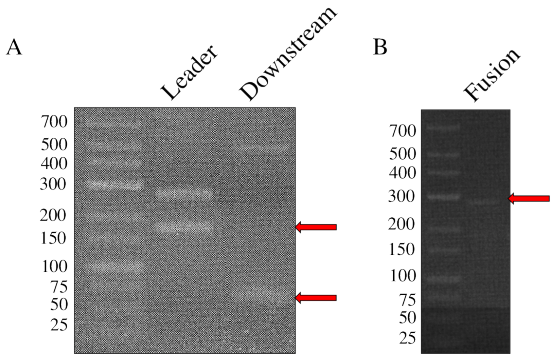


Figure 5.9: A) Cloning of ribosomal RNA leader and downstream regions. For the ribosomal leader sequence, two PCR products were obtained. The lower molecular weight sample was of the correct size. The bands were excised, gel extracted, and subsequently cloned into a pGEM-T vector. B) A splice-by-overlap extension methodology was used to fuse the leader and downstream elements. The PCR product was also gel extracted and ligated into a pGEM-T vector.

gene in archaea (Figure 5.10B).

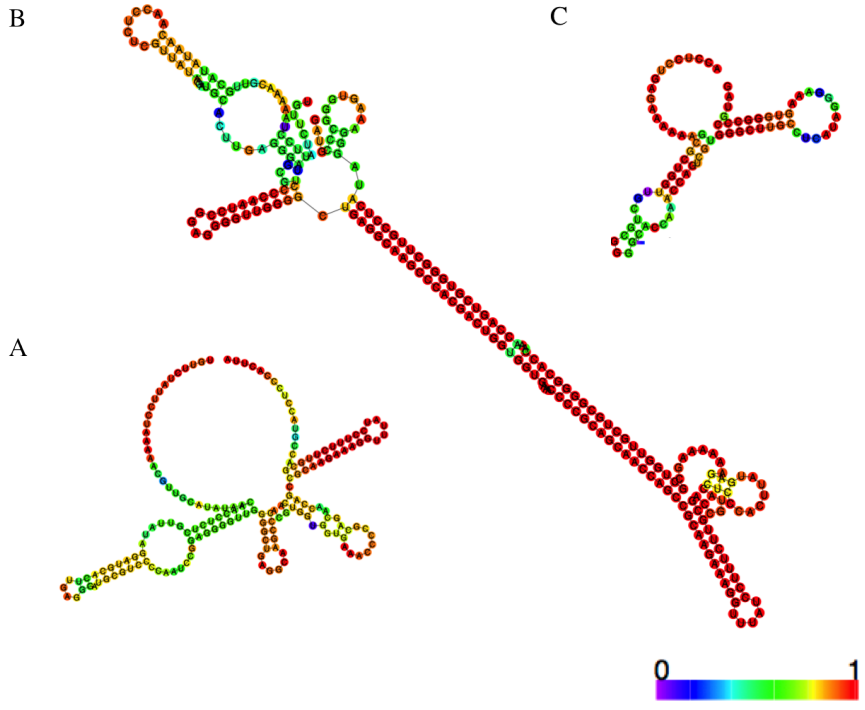


Figure 5.10: Predicted folds of A) 16S rRNA leader, B) leader-downstream fusion, C) downstream regions of the *M. jannaschii rrnA* 16S gene.

5.4.1 EMSAs Using Riboprobes

EMSA assays were performed using the NusA mutants and the riboprobes (Figure 5.11). RNA was at the same concentration in all reactions. 10 μl of riboprobe diluted 1:50 in RNase-free water was added to the reaction in a final volume of 20 μl. Varying final concentrations of protein were used in each reaction (25 nM, 125 nM, 250 nM, 500 nM, 1 μM, 2 μM, and 10 μM). As with the A2 RNA bandshifts, the KH1 GDDG mutant has lower affinity for the leader, fusion, and downstream RNA, whereas the KH2 GDDG mutation does not decrease the affinity. For the wild-type protein, strong complex formation is evident at 500 nM for the leader, fusion, and downstream riboprobes. For the KH1 GDDG mutant strong complex formation is not observed until 1 μM, although there is still a weak shift at 500 nM. The KH2 GDDG mutant shows higher affinity for the leader and downstream RNA than for the fusion. A strong band shift is present at 250 nM for the leader and downstream RNA, as opposed to 500 nM for the fusion. Whilst the general observations of RNA-binding of the mutants is similar to the A2 RNA, the binding is of higher affinity for the riboprobes. For the A2 RNA, complex formation is not evident until the protein concentration reaches the region of 5 μM (Figure 5.5). Band shifts occur at an order of magnitude lower for the riboprobes indicating that NusA binds the ribosomal leader and downstream regions with higher affinity than the A2 RNA. NusA-bound RNA was quantified as a percentage of total RNA from a control lane containing no NusA using ImageJ.

5.5 Scaffold-Independent Analysis

In order to gain insight into the possible function of archaeal NusA, a technique called scaffold-independent analysis (SIA) (Beuth et al., 2007) was performed to determine what, if any, preference NusA has for RNA sequence. SIA is an HSQC-based assay that enables determination of the sequence specificity of an RNA-binding KH domain. It can determine positive and negative discrimination of nucleotides. In order to perform SIA, the protein of interest is incubated with pools of RNA pentamers. This is because KH domains bind a region of RNA between 4–5 nucleotides long (Lewis et al., 2000). The first nucleotide in the pentamer is bound at its $2'$ OH and phosphate group in a non-sequence-specific manner (Lewis et al., 2000). Therefore in SIA the first nucleotide of the pentamer is always a U and referred to as position 0. The other

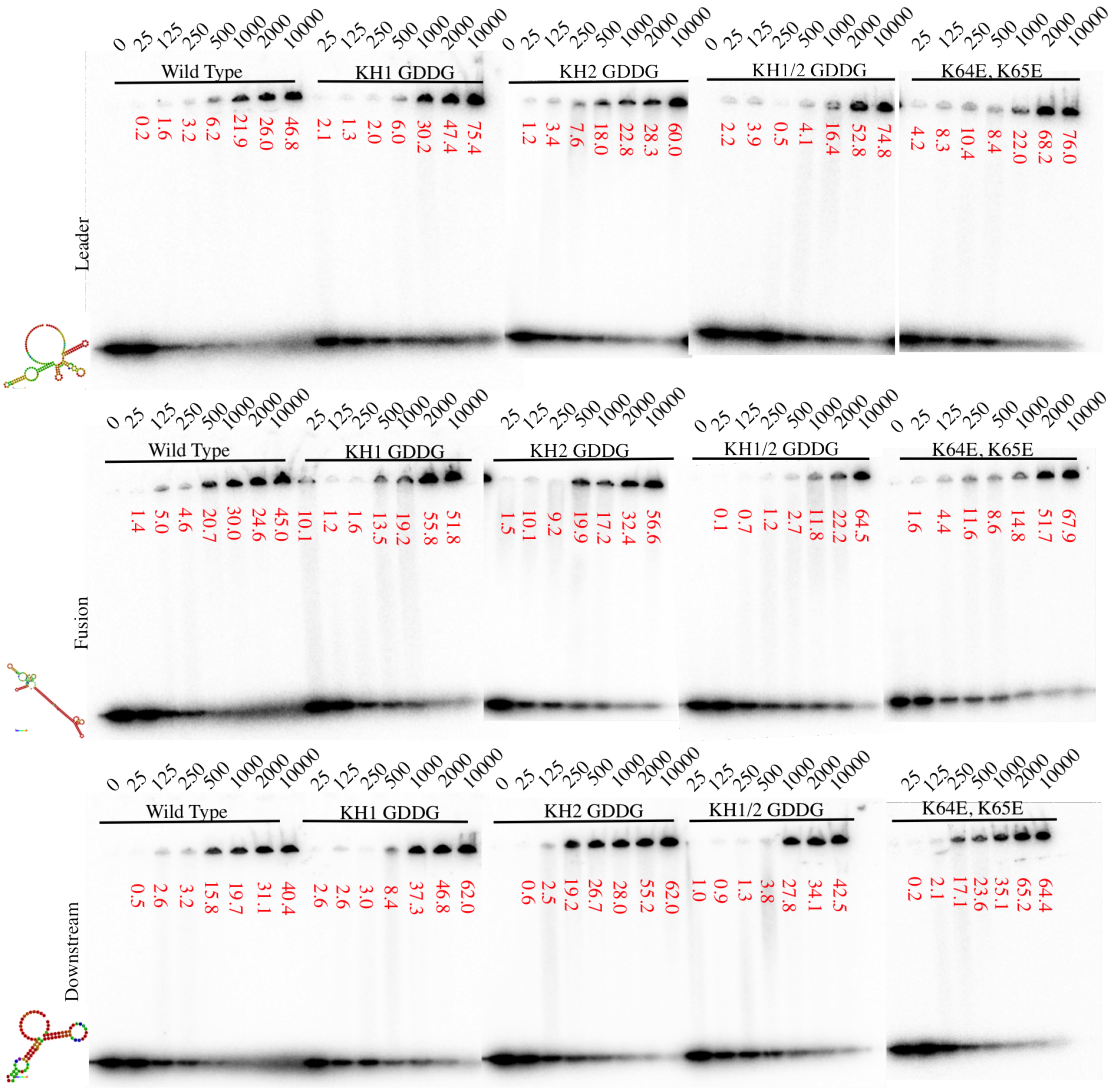


Figure 5.11: EMSA performed with the NusA mutants on the rRNA(A) 16S leader, and rRNA(A) 16S fusion riboprobes. Concentration of protein in nM shown above wells. NusA-RNA complexes were quantified using ImageJ (red).

four nucleotides, however, are bound in a sequence-specific manner. Three out of the remaining four positions are random (either A, U, G or C, referred to as N). In the remaining position the identity of the nucleotide is known and always constant in a given RNA pool. The preference of the KH domain for each position of the RNA pentamer is analysed independently of the other positions. Thus for each position, four HSQC spectra are recorded (one for each of the 4 RNA nucleotides). Each spectrum is recorded with a pool of 64 RNA pentamers (There are three random nucleotides within each pentamer, each of which can take one of four identities. This

	Position 1	Position 2	Position 3	Position 4
A pools	UANNN	UNANN	UNNAN	UNNNA
C pools	UCNNN	UNCNN	UNNCN	UNNNC
G pools	UGNNN	UNGNN	UNNGN	UNNNG
U pools	UUNNN	UNUNN	UNNUN	UNNNU

Table 5.1: RNA oligo pools used in SIA. For each of the four analysed positions, four separate HSQC spectra are recorded in which the identity of the nucleotide at the analysed position is known and constant (one of A, U, G, or C). The nucleotide identity at the remaining three positions is random (N). Thus in each library there is a pool of 64 different RNA pentamers. The nucleotide preference at each position is determined independently of the other three positions.

gives 4^3 possible variants.) (Table 5.1). One advantage of SIA analysis is that it allows the determination of RNA-binding specificity under equilibrium conditions. RNAs with a higher affinity for the protein will have a greater proportion bound, resulting in a larger change in the chemical shifts of the HSQC spectra. The relative bias for a nucleotide at any given position is ascertained by comparing the perturbation size of the change in the chemical shift relative to the spectrum of free protein.

5.5.1 HSQC NMR Spectroscopy

Some atomic nuclei have spin. Spin has already been discussed with respect to electrons (see Chapter 4). ^{1}H, ^{15}N, and ^{13}C are commonly used in protein NMR. Since ^{15}N and ^{13}C are not naturally abundant, they must be incorporated into proteins by expressing them in minimal media where the nitrogen and carbon sources are highly enriched in the desired isotope (see Chapter 2). Like electrons, ^{1}H, ^{15}N, and ^{13}C have a spin number of 1/2. This means that when placed in an external magnetic field they can take one of two possible orientations. In an analogous manner to EPR, atomic nuclei that have spin can be excited from the lower to the higher energy state by providing energy that is equal to the difference between the two spin states. Atomic nuclei resonate in the MHz region, compared to electrons which resonate in the GHz region. It is possible for magnetisation to be transferred between nuclei. In ^{15}N-^{1}H HSQC spectroscopy magnetisation is transferred from the ^{1}H that is bonded to the ^{15}N, to the ^{15}N, and then back to the ^{1}H. The ^{1}H and ^{15}N chemical shifts are therefore correlated. This results in a 2D spectrum where all NH groups give rise to a peak. All amino acids are represented by a backbone NH peak except for the imino acid proline which has no NH group. In addition the NH of tryptophan, asparagine, and glutamine side chains also give rise to shifts. An HSQC spectrum can pro-

vide information regarding whether a protein is well folded, in which case the spectrum will be well dispersed. The position of the chemical shifts may be affected by interaction with another biological molecule, making HSQC a useful technique for probing interactions between molecules.

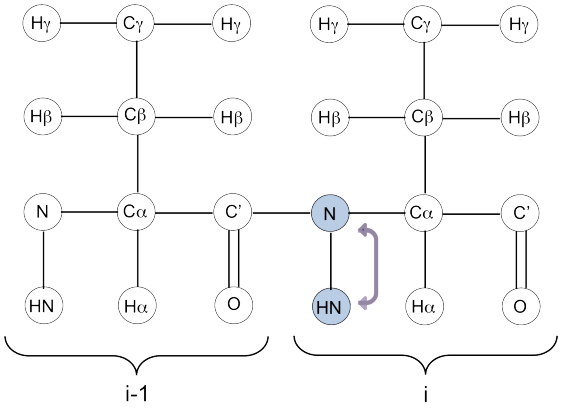


Figure 5.12: Transfer of magnetisation in an HSQC experiment. Magnetisation is transferred from the ^{1}H that is bonded to the ^{15}N, to the ^{15}N, and then back to the ^{1}H.

SIA only works on single KH domains. Since NusA contains two KH domains it was necessary to selectively knock out their RNA-binding activity. Otherwise both KH domains would bind the RNA pentamers, resulting in two signals which could not be separated. Therefore the KH1 GDDG mutant was used.

5.5.2 The NusA KH2 Domain Binds RNA with Little Preference for Sequence

Initially the HSQC spectrum of the free protein was overlaid against a spectrum containing a completely random pentamer to determine whether RNA addition resulted in crosspeak shifts. Once this was established, the SIA analysis was carried out. All HSQC spectra were recorded at a protein concentration of 70 μM. RNA pentamers were also added at 70 μM. The spectra were recorded on a *Brucker* Avance spectrometer, operating at 700 MHz.

Each RNA-binding spectrum was overlaid with the spectrum of free protein (Figure 5.14) and the peaks that changed upon RNA binding were identified. Eight chemical shifts were selected

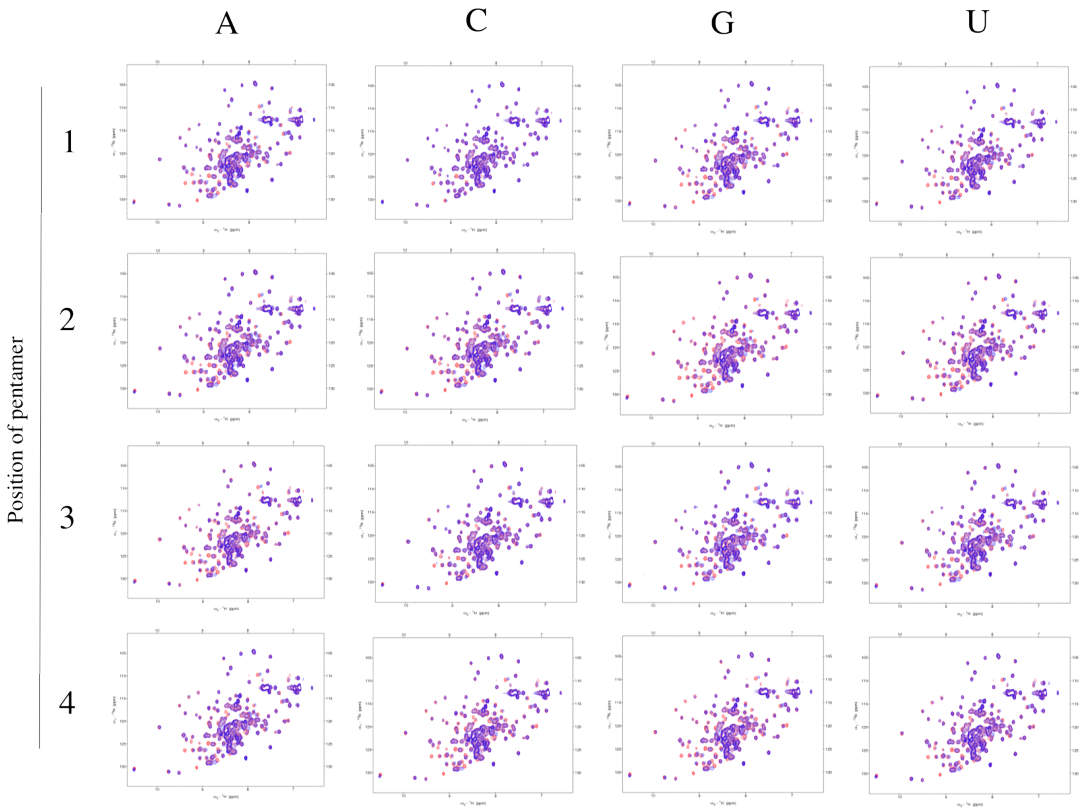
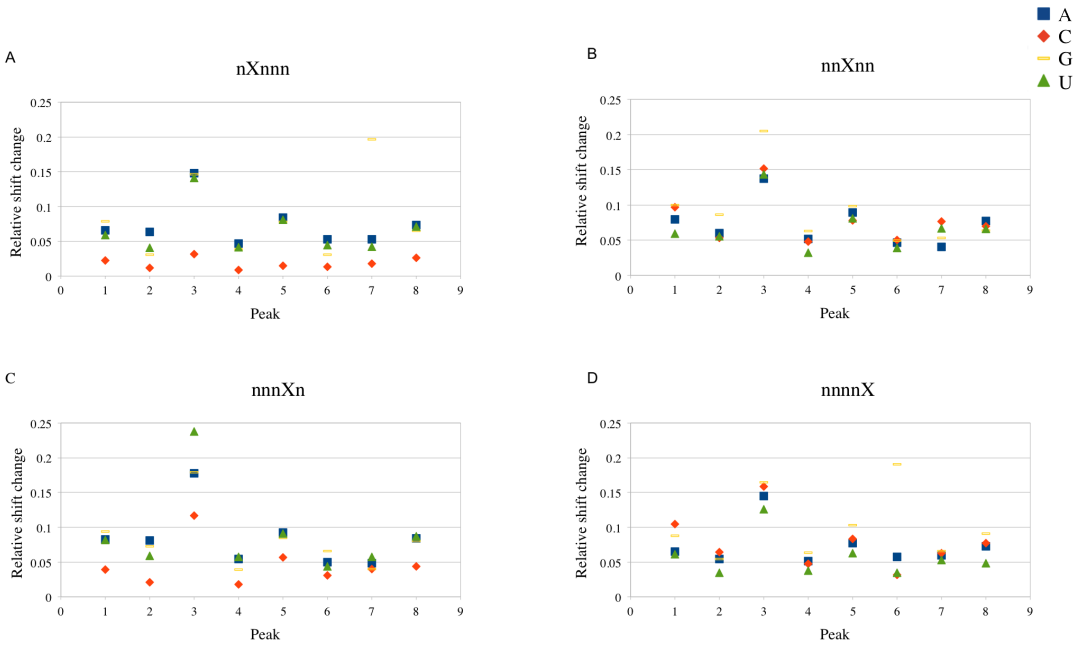


Figure 5.13: HSQC spectra obtained in NusA KH1 GDDG SIA experiment showing A, C, G, and U pools for positions 1, 2, 3, and 4. Spectra of free protein in red. Spectra of samples containing a 1:1 ratio of protein:RNA are overlaid in blue.

that changed in each spectrum and each position of the pentamer was analysed independently. For each of the eight selected shifts, four relative chemical shift changes were determined upon RNA addition (one for each of the four nucleotides). This typically gave values in the range of 0–0.3 ppm. For each of the eight selected peaks in the spectra, the four relative shift changes were normalised to the highest value by dividing them by the highest relative distance change. This gave a number between 0–1 for each possible nucleotide at each analysed position of the pentamer. The eight normalised shift changes were averaged to give the final SIA selectivity scores. This was repeated for each of the four analysed positions of the pentamer, and the allowed and disallowed nucleotides were determined.

SIA on the KH2 domain of NusA did not reveal strong binding specificity, although there does appear to be negative selection for particular nucleotides in positions 1, 3 and 4. In position 1,



E

Position	A	C	G	U
1	0.89	0.22	0.84	0.78
2	0.79	0.85	0.94	0.72
3	0.89	0.47	0.86	0.91
4	0.74	0.81	0.96	0.58

Figure 5.14: Relative shift changes for the eight selected peaks upon RNA addition at A) position 1, B) position 2, C) position 3, and D) position 4. For each peak A is blue, C is orange, G is yellow, and U is green. E) Averaged SIA selectivity scores for each analysed position of the RNA pentamer. The relative affinities for each nucleotide are indicated. Positions 1 and 3 show negative selectivity for C. Position 4 shows negative selectivity for U.

A, G and U are all accommodated but there is negative selection against C. In position 2 all ribonucleotides are accommodated, although G seems to be particularly favoured. In position 3, C is again selected against. In position 4 all nucleotides seem to be bound by NusA reasonably well. G is particularly favoured and there is possibly negative selection against U.

5.5.3 HSQC Analysis of Additional NusA RNA-Binding Mutants

The SIA experiment on the KH1 GXXG–to–GDDG mutant did not demonstrate strong sequence specificity. Given that bandshift assays indicated that NusA may bind RNA via a novel site, HSQC spectra of the NusA RNA-binding mutants were recorded to determine whether it was worthwhile performing SIA on the KH2 GDDG mutant. These spectra were recorded on a *Brucker* Avance, operating at 600 MHz.

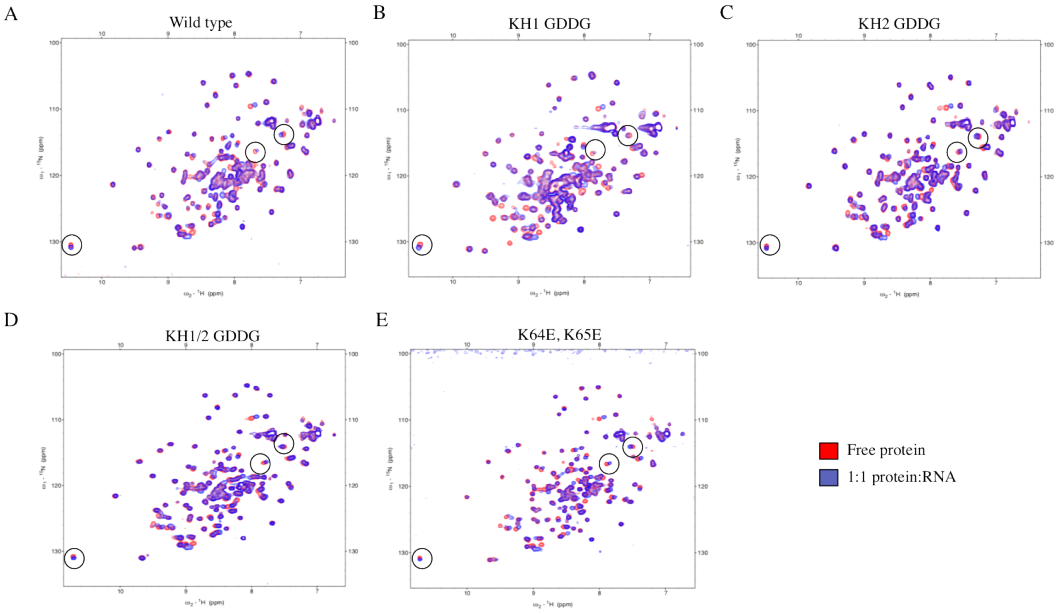


Figure 5.15: HSQC spectra of NusA RNA-binding mutants. Red: free protein. Blue: 1:1 ratio protein-to-RNA. A) WT, B) KH1 GDDG, C) KH2 GDDG, D) KH1/2 GDDG, E) K64E K65E. All RNA-containing spectra recorded with NNNGN pentamers, except KH1 GDDG which was recorded with NNNNC pentamers. Protein and RNA at 1:1 ratio (70 μM). The same shifts change upon the addition of RNA for all mutants.

All of the mutants result in spectra that are well dispersed (Figure 5.15), indicating that they are properly folded and that there is no aggregation. They confirm that the mutants, including the double KH1/2 GDDG mutant, are able to bind to RNA as evidenced by the change in some of the chemical shifts upon the addition of RNA. The EMSAs suggest that the KH1 domain, and not the KH2 domain contributes more strongly to RNA binding, however the HSQC spectra clearly indicate that the same shifts are changing in both the KH1 and the KH2 mutants. Furthermore the same shifts change upon the addition of RNA in all of the mutants, indicated in black circles. This is an unexpected observation and strongly suggests that NusA binds RNA via a novel binding site, using residues other than the GXXG motifs and the K64, K65 region. This also indicates that performing SIA on the KH2 GDDG mutant may not be informative because SIA only works for canonical KH domains.

5.6 Analysing NusA by 3D NMR

5.6.1 Expression and Purification of Double-Isotopically Labelled NusA

Three-dimensional NMR was performed on NusA to determine precisely which residues are involved in the binding of RNA. Three-dimensional NMR requires protein that has been isotopically labelled with ^{15}N and ^{13}C. In addition the protein was expressed in D_2O media. Deuterated proteins have a decreased rate of relaxation of their spins from dipole-dipole interactions and therefore give spectra of greater quality. Although about 80% of the hydrogens within the protein are ^{2}H, the NH hydrogens are able to exchange with the solvent. Therefore ^{1}H resonances can still be observed on proteins expressed in D_2O media. An overnight culture was grown in minimal media and used to inoculate 3 L the following morning. Protein expression was induced at an OD_{600} of 0.6, and the cells were harvested 5 hours thereafter. NusA was extracted, purified on a heparin column and concentrated to approximately 400 μM. The sample was analysed by NMR using an HNCACB pulsed experiment. Within this experiment resonance is transferred from the ^{1}Hβ to ^{13}Cβ, and then from ^{13}Cβ to ^{13}Cα. Resonance is also simultaneously transferred from ^{1}Hα to ^{13}Cα. From ^{13}Cα the resonance is transferred to ^{15}N, and finally to the ^{1}H of the backbone amide (Figure 5.16A). Thus there are three dimensions to the resulting spectrum: ^{13}C, ^{15}N, and ^{1}H. HN(CO)CACB experiments were also conducted. In the HN(CO)CACB experiment, resonance is transferred from ^{1}Hβ to ^{13}Cβ, and then from ^{13}Cβ to ^{13}Cα. Resonance is also simultaneously transferred from ^{1}Hα to ^{13}Cα. It is then transferred to ^{13}CO, and then to the ^{15}N of the backbone amide and finally to the ^{15}NH (Figure 5.16B). Again, there are three dimensions to the spectrum. These two experiments are both routinely used together to assign a protein backbone in an NMR spectrum. Together, they enable shifts from one residue to be linked to shifts from the next residue, which enables sequential assignment of residues.

5.6.2 Assigning the NusA Residues that Bind RNA

The NMR spectrum is currently being assigned by Katherine Collins of Professor Andres Ramos's group and is approximately 60% complete. It is already clear that the residues currently as-

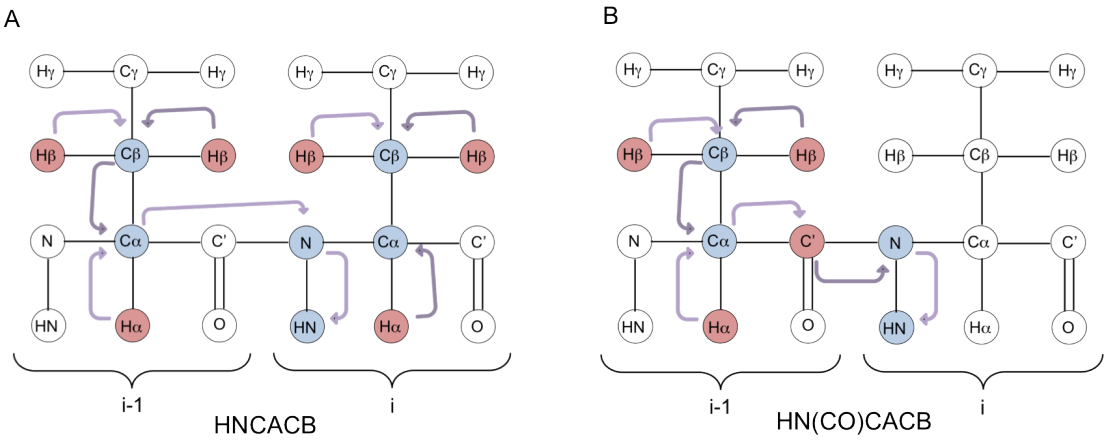


Figure 5.16: Transfer of magnetisation in 3D NMR experiments used to assign the backbone of *M. jannaschii* NusA. A) In the HNCACB experiment magnetisation is transferred from $^{1}H\beta$ to $^{13}C\beta$. Then it is transferred from $^{13}C\beta$ and $^{1}H\alpha$ to $^{13}C\alpha$. Finally magnetisation is transferred to ^{15}N, and then to the ^{1}H of the backbone amide. B) In the HN(CO)CACB experiment magnetisation is transferred from $^{1}H\beta$ to $^{13}C\beta$. Then it is transferred from $^{13}C\beta$ and $^{1}H\alpha$ to $^{13}C\alpha$. From here it is transferred to ^{13}CO, and then to the ^{15}N of the backbone amide and finally to the ^{15}NH.

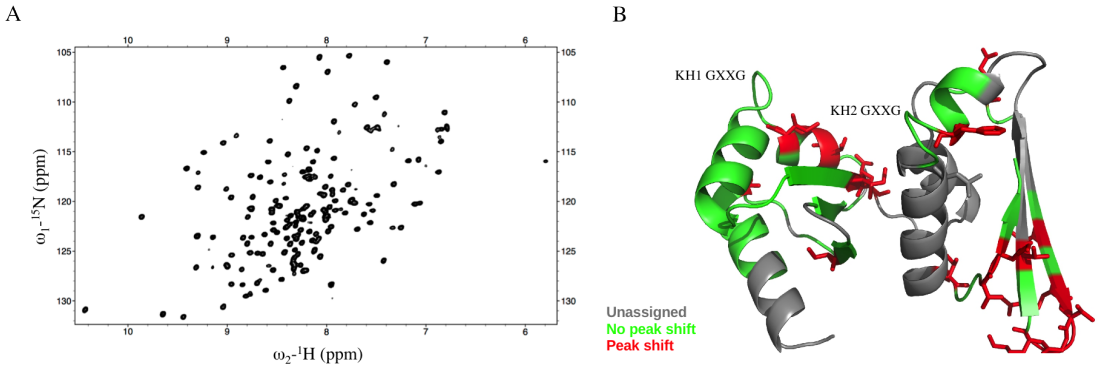


Figure 5.17: A) 3D NMR spectrum of *M. jannaschii* NusA and current assignment of residues that bind RNA. A) 3D Spectrum showing ^{15}N and ^{1}H dimensions. B) Homology model of *M. jannaschii* NusA showing currently assigned residues (green). Residues predicted to bind RNA in red.

signed as RNA-binding are not similar to those in bacteria (Figure 5.17). The residues assigned as binding RNA are across the entire surface of the homology model, unlike in bacteria where the RNA spans the two GXXG loops only. In contrast, the two GXXG loops in archaea, both of which have been assigned, do not shift upon the addition of RNA. This suggests either that *M. jannaschii* NusA binds to RNA via a radically different mechanism to bacterial NusA, or that the assignment is not providing an accurate model of how NusA actually binds RNA *in*

vivo. Support for the assignment comes from the ^{15}N–^{1}H HSQC spectra that were performed on the RNA-binding mutants. These indicated that the same shifts change upon RNA addition for all mutants. This in turn strongly suggests that residues other than the GXXG motifs are binding the RNA. One factor which could be problematic is the use of RNA pentamers. Whilst individual KH domains bind regions of RNA 4–5 nucleotides long, the region of RNA that NusA binds *in vivo* is part of a long transcript, not an isolated pentamer. Using longer RNAs in an NMR experiment may not be feasible, however. This may enable multiple copies of NusA to bind to individual RNAs, which would in turn decrease the rate of tumbling and result in poor dispersion of the NMR spectrum.

5.7 Discussion

Within this chapter the RNA-binding properties of *M. jannaschii* NusA have been analysed. Bandshift assays were performed to determine regions of NusA that bind to RNA. Mutation of the KH1 GXXG loop impairs RNA binding, but mutation of the KH2 GXXG loop has no effect. Combining the two GXXG–to–GDDG mutations with the K64E, K65E mutation completely abolishes RNA binding, indicating a role for the two lysine residues in the binding of RNA. This reveals that NusA may be interacting with RNA via a novel binding site. Comparison of bandshift assays using A2 RNA and the 16S leader/downstream regions suggests that NusA does bind some sequences preferentially to others. Complex formation occurs at a 10-fold lower NusA concentration for the biologically relevant riboprobes than for the A2 RNA.

The SIA analysis of the KH1 GXXG–to–GDDG mutant did not reveal a strong sequence bias for NusA. The lack of apparent sequence specificity may indicate that NusA does not have a sequence-specific function in archaea. However, the data also points to NusA associating with RNA via a novel binding site which has not been observed in KH domain proteins before. This explains why the SIA did not provide a clear sequence specificity; SIA analyses canonical KH domains, but *M. jannaschii* NusA does not bind RNA in a canonical fashion.

Analysis of the remaining KH domain mutants by HSQC indicates that the same peaks change in their chemical shifts upon the addition of RNA for the KH1 and KH2 mutants. This provides

yet further evidence that NusA is binding RNA via a novel binding site, and utilising regions of the protein other than the GXXG loops and the lysine residues K64 and K65. 3D NMR was performed to allow assignment of all residues that are involved in the binding of NusA. Currently the spectrum is approximately 60% assigned. From the peaks that have been assigned it is clear that NusA is interacting with the RNA in an unexpected way. The two GXXG loops are apparently not involved at all. Those residues whose chemical shifts change upon the addition of RNA are scattered across the protein, and there is no obvious region or path where the RNA is binding. This may be due to the fact that RNA pentamers were used in the experiment, as opposed to more biologically relevant longer RNAs. Upon completion of the assignment it will be possible to reach a clearer conclusion as to whether RNA really does bind across all over the NusA surface, or whether a new approach needs to be taken to successfully create a model of RNA binding.

An alternative approach to determine the sequence specificity of NusA would be to use RNA SELEX (Systematic Evolution of Ligands by Exponential Enrichment). SELEX enables the *in vitro* selection of nucleic acid aptamers that are bound by a protein (Stoltenburg et al., 2007). It requires the production of a large library of random DNA sequences with $5'$ and $3'$ adaptors that are targets for primers. The DNA library must be converted to RNA and is then incubated with the immobilised protein of interest. Aptamers that do not bind are washed away. The aptamers that do bind are reverse-transcribed and amplified by PCR using primers against $5'$ and $3'$ adaptors. The enriched aptamers are re-transcribed into RNA and the cycle is repeated. The binding conditions increase in stringency each round to select for the highest affinity binders. The sequences of the aptamers can then be determined and the sequence preference of the protein identified. SELEX has the advantage that large RNAs can be analysed. This has the additional advantage of allowing selection for and against RNA secondary structure, which is a very important factor in protein-RNA interactions.

Chapter 6

Conclusions and Perspectives

6.1 The State of the Art

RNAP is an essential multi-subunit enzyme found in all domains of life that carries out the process of transcription (Werner and Grohmann, 2011). RNAPs across the three domains of life show a high degree of structural conservation with a characteristic 'crab claw' structure. The archaeal transcription machinery represents a streamlined version of that found in eukaryotes. Archaeal RNAP closely resembles eukaryotic RNAPII and contains a number of subunits that are not found in bacteria, most notably the Rpo4/7 stalk.

RNAP is a highly processive enzyme that is able to transcribe 1000s of nucleotides at a time. However, RNAP is also subject to frequent pausing. Pausing provides opportunity for regulation by facilitating the binding of transcription factors to the elongation complex, but it also slows transcription (Landick, 2009). Various general transcription elongation factors function to reduce pausing and promote processive elongation. The Gre factors in bacteria and TF(II)S in archaea/eukaryotes promote RNA cleavage in backtracked RNAP. The result of this cleavage is that the 3′ end of the RNA is correctly positioned in the active site for incorporation of further nucleotides. The Gre factors and TF(II)S act via the same mechanism but are not homologous (Kettenberger et al., 2003; Opalka et al., 2003). NusG and its archaeal/eukaryotic homologue Spt5, in contrast, are the only universally conserved RNAP-associated transcription elongation factors (Werner and Grohmann, 2011). NusG contains an N-terminal NGN domain and a C-terminal KOW domain. It increases the processivity of RNAP by binding to the coiled-coil of the RNAP clamp domain, thereby enclosing the DNA entry channel. Spt5 in archaea and eukaryotes functions via the same mechanism (Hirtreiter et al., 2010a). Whilst the processivity-enhancing function of NusG proteins is conserved across the three domains of life, NusG proteins have evolved additional functions that differ between bacteria, archaea, and eukaryotes. These additional functions are mediated via the KOW domain(s).

In bacteria the multiple functions of NusG are well-documented. The KOW domain binds to the termination factor rho (Li et al., 1993), facilitating the silencing of foreign DNA (Cardinale et al., 2008). The NusG KOW domain is also able to bind to the ribosomal protein S10, suggesting that it is able to directly couple transcription and translation by bridging RNAP and

the ribosome (Burmann et al., 2010b). In eukaryotes, the functions of Spt4/5 have also been well-documented. In metazoans, Spt4/5 is required to mediate the phenomenon of promoter-proximal stalling, where RNAP pauses about 40 nucleotides downstream of the transcription start site in an activated state (Rougvie et al., 1988). The Spt5 KOW domains also bind RNA processing and nucleosome remodelling factors (Lindstrom et al., 2003).

The functions of Spt4/5 in archaea are less well characterised. Whilst it is known that the NGN domain of Spt5 binds to RNAP, and that Spt4/5 increases the processivity of RNAP, very little is known about its other functions. For example, it is not known whether Spt5 is capable of binding to ribosomes. Earlier data suggests that Spt4 stabilises the NGN domain but this has never been quantified (Hirtreiter et al., 2010a). *T. maritima* NusG, and RfaH both exhibit conformational flexibility between their N-terminal and C-terminal domains (Belogurov et al., 2007; Drögemüller et al., 2013). Archaeal Spt4/5 has not been examined to determine whether there is flexibility between the NGN and KOW domains. However, the fact that it has not been possible to crystallise full length *M. jannaschii* Spt4/5 suggests that there are flexible regions within the protein.

NusA is a protein that also binds to RNAP thereby modulating transcription (Schmidt and Chamberlin, 1987). Unlike NusG, it is not universally conserved and is present in bacteria and archaea only. In bacteria, NusA promotes pausing when acting in isolation. However, NusA also functions as a component of the antitermination complex, in which it has the opposite function (DeVito and Das, 1994). The antitermination complex enables RNAP to enter a pause and termination-resistant state. This protects against rho-dependent termination, and also enables RNAP to read through intrinsic terminators. In *E. coli*, antitermination was discovered in phage λ-infected cells where the phage protein N hijacks the host proteins NusA, NusB, NusE (S10) and NusG to promote expression of its genome. NusA also functions to promote the correct and efficient folding of rRNA (Bubunenko et al., 2013).

The functions of NusA in archaea are also less well characterised. Archaeal NusA is a streamlined version of that found in bacteria. It contains two KH domains only. It is not known therefore whether it retains its RNAP-binding activities. It has previously been shown that NusA binds

to RNA, but no attempt has been made to determine its sequence specificity or to dissect the relative contributions of the NusA domains to RNA binding (Shibata et al., 2007).

6.2 Contribution to Knowledge Within This Thesis

This thesis presents an in-depth biochemical and biophysical analysis of Spt4/5 from the archaeon *M. jannaschii*. Archaeal cell extract was separated by Superose 6 size exclusion chromatography. The elution profiles of the ribosome (S10) and RNAP (Rpo4/7) were characterised. It was then shown that Spt4/5 elutes in three peaks. These are ribosome-bound, RNAP-bound, and free protein. This provides evidence, for the first time, that archaeal Spt4/5 may have a role in translation. The Spt4/5 complexes are nuclease-sensitive, with Spt4/5 dissociating from the transcription and translation complexes upon DNase treatment. This is consistent with Spt4/5 being a transcription factor that binds to actively transcribing DNA-bound RNAP. The interaction between Spt4/5 and the ribosomes was then analysed using MST. This gave a K_d in the low μM range and is the first time the interaction between Spt4/5 and whole ribosomes has been quantified. The affinity of the interaction is consistent with transcription being a dynamic process. The interaction needs to be strong enough to allow robust gene expression. However, dissociation of Spt4/5 from both RNAP and the ribosome is required for transcription termination and for repetition of the transcription cycle.

DEER was performed on double nitroxide spin-labelled Spt4/5. Spt4/5 exhibits flexibility between the NGN and KOW domains of Spt5. The data suggests that Spt5 is able to exist in more than one conformation and that the presence of Spt4 restricts its mobility. This is the first time that distinct Spt5 conformations have been observed in archaea. Spt4 may promote the expression of distinct gene subsets or convert Spt5 from a processivity-enhancing factor to a transcription-translation coupling factor. This could be achieved through Spt4 positioning the KOW domain, allowing for the recruitment of transcription factors. Thermofluor assays on Spt4/5 domain-deletion mutants demonstrated that Spt4 stabilises the NGN domain of Spt5. The NGN domain does not denature at 98°C when it is bound by Spt4. The NGN domain denatures at 85°C when unbound. This is the first time that the stabilisation of the NGN domain by Spt4 has been quantified. Limited proteolysis assays demonstrated that Spt5 is more susceptible

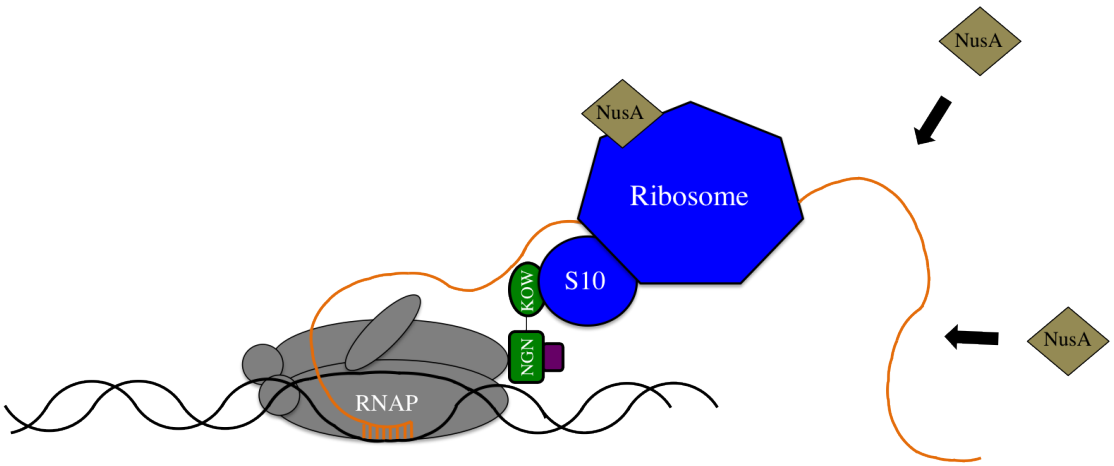


Figure 6.1: Summary of factor interactions determined within this thesis. The KOW domain of Spt5 binds to the S10 protein of the ribosome. NusA is stably associated with the ribosome and binds RNA with little sequence specificity.

to digestion when Spt4 is removed. Spt5 is cleaved into two fragments that likely correspond to the NGN and KOW domains. This suggests that there is flexibility in the inter-domain linker and therefore supports the data obtained by DEER. Spt4 restricts the flexibility between the two domains of Spt5 by making contacts with the flexible linker (Figure 6.2).

NusA from fractionated cell extract was also shown to elute in the same fractions as the ribosome, but not in the same fractions as RNAP. This is consistent with the fact that the bacterial NusA N-terminal domain, which binds RNAP, is not present in archaea. Analysis of the size fractionation of NusA from non-nuclease-treated cell extract indicated that NusA elutes in a broad peak. This suggests that it is bound to multiple RNAs of various sizes, rather than to a specific species of RNA. Analysis of purified ribosomes on a Superose 12 column confirmed that NusA is stably associated with the ribosome.

The RNA-binding of *M. jannaschii* NusA was analysed. A homology model of *M. jannaschii* NusA was generated, and this aligned well with the structure from *A. pernix*. Band shift assays on KH domain GXXG–to–GDDG mutants indicated that the KH1 domain of NusA contributes to RNA-binding much more than the KH2 domain. The bandshift assays also suggested that NusA may bind RNA via a novel binding site, as NusA was still able to bind to RNA when both GXXG loops had been mutated. Combining the two KH GXXG–to–GDDG mutations with the

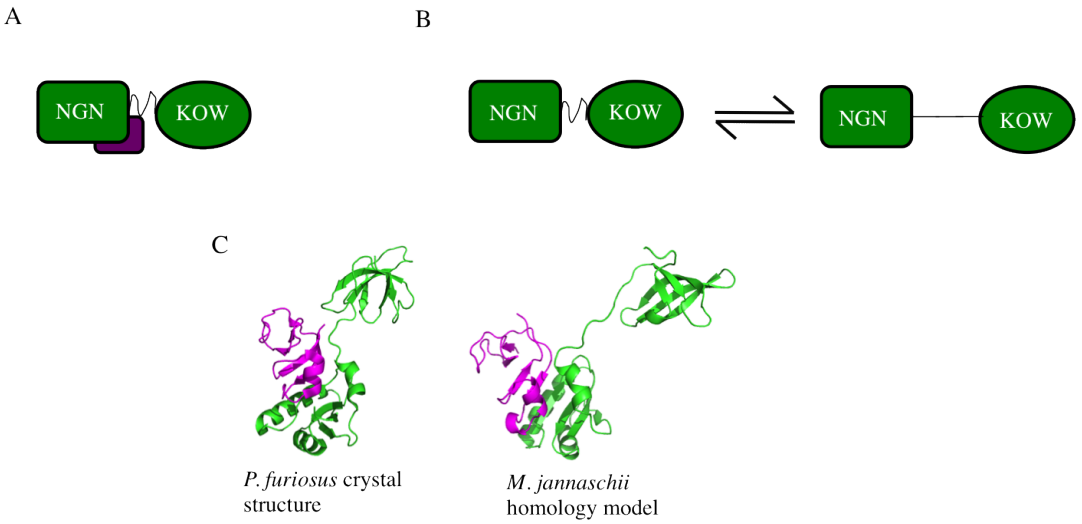


Figure 6.2: Summary of Spt5 conformations. A) Spt4 binds to the NGN domain and protects the flexible linker. This also restricts the conformational flexibility of Spt5. B) In the absence os Spt4, there is more conformational flexibility in the Spt5 interdomain linker. C) Comparison of the *P. furiosus* Spt4/5 crystal structure and the *M. jannaschii* homology model. In the *P. furiosus* crystal structure, Spt4 is observed to make contact with the flexible linker. Spt5 is more compact in the crystal structure. In the homology model Spt4 does not protect the flexible linker and Spt5 is more elongated.

charge reversal of two lysine residues (K64 and K65) to glutamate completely abolished binding. HSQC analysis supports the notion that NusA binds RNA via a novel binding site, as the same peaks change their chemical shifts upon the addition of RNA in all of the mutants. Repeating the EMSAs using the *rrnA* leader and downstream regions, and a fusion of the two resulted in complex formation at lower concentrations of NusA. For the Cy3-labelled A2 RNA, bandshifts were observed at 5 μM NusA whereas for the rRNA riboprobe, bandshifts were observed at 250–500 nM. This represents a 10-fold higher affinity suggesting that NusA does have some sequence bias. However SIA analysis on the KH1 GXXG–to–GDDG mutant did not yield a clear sequence bias. It is therefore not clear whether archaeal NusA preferably binds to the *rrnA* leader and downstream regions *in vivo*. ^{13}C, ^{15}N, ^{2}H labelled NusA was prepared and a 3D NMR spectrum was recorded to enable assignment of residues that are involved in the binding of NusA. Approximately 60% of the protein has currently been assigned and will hopefully shed new insights into the mechanism of RNA binding. However it is already clear that something interesting is occurring. The GXXG motifs have been identified as non-RNA binding. GXXG motifs are usually considered the key players in KH domain-RNA interactions. It will hopefully become

clearer how *M. jannaschii* NusA binds RNA upon completion of the NMR spectrum assignment.

6.3 Future Perspectives

This thesis has provided significant insights into the function of the archaeal Spt4/5 and NusA proteins. However, these insights raise further questions regarding how these factors carry out their roles. Superose 6 separation of cell extract and MST indicate that Spt4/5 interacts with ribosomes. The interaction could be further confirmed by co-pelleting experiments, which have the advantage of not requiring ribosome labelling. The functional interplay between RNAP, Spt4/5, and the ribosome could be analysed by developing an *in vitro* coupled transcription-translation system, similar to that which has been reported in bacteria (Castro-Roa and Zenkin, 2012).

In vitro studies enable the molecular mechanisms of complex processes to be elucidated. However, the true biological significance of these studies can only be ascertained through *in vivo* validation. There is no genetics system available for *M. jannaschii*. However, it is possible to knock out proteins and to introduce mutants into the crenarchaea *Sulfolobus acidocaldarius* (Wagner et al., 2009), which our lab has recently started studying. Introducing mutations into Spt4/5 that abolish its interaction with the ribosome and analysing the effects on gene expression would indicate whether the interaction between Spt4/5 and the ribosome regulates transcription.

Data within this thesis demonstrates that Spt4 stabilises the NGN domain of Spt5. It would be interesting to determine whether this stabilisation is essential for cell viability by attempting to knock out Spt4. Analysis of global transcription in the presence and absence of Spt4 would identify whether Spt4 is responsible for regulation of specific subsets of genes.

The Superose 6 size separation of *M. jannaschii* cell extract provides strong evidence that NusA does not bind to RNAP. Furthermore the ribosomal protein S10 is only found in ribosome-associated fractions and not in its free form. This strongly suggests that NusA and S10 do not not participate in an antitermination complex. Further insights into the role of NusA could also be obtained by taking an *in vivo* approach and determining where within the transcrip-

tome it binds. For example, does NusA bind to the *rrn* operon leader sequence and regulate ribosome biogenesis? This could be determined by using a method such as CLIP which enables determination of the *in vivo* occupancy of RNA-binding proteins (Ule et al., 2005). In CLIP, proteins are covalently cross-linked to RNA using UV light and the cells are then disrupted. The protein-RNA complexes are subsequently immunoprecipitated, and the RNA is partially digested then radiolabelled. After SDS PAGE analysis the cross-linked protein is transferred to a nitrocellulose membrane and excised. The protein is digested by proteinase K, and the RNA is reverse-transcribed to cDNA, amplified by PCR, and sequenced. The sequences can then be identified in the transcriptome.

Another line of study would be to introduce the NusA RNA-binding mutations into archaeal cells and analyse the cell viability. It would be interesting to determine whether one of the KH domains has a more important role *in vivo*, and whether this would validate the *in vitro* EMSAs presented in thesis.

Bibliography

Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R., and Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438(7067):460–465, 2005.

Adelman, K., Kennedy, M. A., Nechaev, S., Gilchrist, D. A., Muse, G. W., Chinenov, Y., and Rogatsky, I. Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proceedings of the National Academy of Sciences*, 106 (43):18207–18212, 2009.

Albrechtsen, B., Squires, C. L., Li, S., and Squires, C. Antitermination of Characterized Transcriptional Terminators by the *Escherichia coli rrnG* Leader Region. *Journal of molecular biology*, 213(1):123–134, 1990.

Altenbach, C., Flitsch, S. L., Khorana, H. G., and Hubbell, W. L. Structural Studies on Transmembrane Proteins. 2. Spin Labeling of Bacteriorhodopsin Mutants at Unique Cysteines. *Biochemistry*, 28(19):7808–7812, 1989.

Altenbach, C., Marti, T., Khorana, H. G., and Hubbell, W. L. Transmembrane Protein Structure: Spin Labeling of Bacteriorhodopsin Mutants. *Science*, 248(4959):1088–1092, 1990.

Armache, J.-P., Anger, A. M., Márquez, V., Franckenberg, S., Fröhlich, T., Villa, E., Berninghausen, O., Thomm, M., Arnold, G. J., Beckmann, R., and Wilson, D. N. Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Research*, 41(2):1284–1293, 2013.

Armache, K.-J., Mitterweger, S., Meinhart, A., and Cramer, P. Structures of Complete RNA Polymerase II and Its Subcomplex, Rpb4/7. *Journal of Biological Chemistry*, 280(8):7131–7134, 2005.

Artsimovitch, I. and Landick, R. The Transcriptional Regulator RfaH Stimulates RNA Chain Synthesis after Recruitment to Elongation Complexes by the Exposed Nontemplate DNA Strand. *Cell*, 110(6):193–203, 2002.

Bailey, M. J. A., Hughes, C., and Koronakis, V. RfaH and the *ops* element, components of a novel system controlling bacterial transcription elongation. *Molecular Microbiology*, 26(5): 845–851, 1997.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A., and Nudler, E. A Ratchet Mechanism of Transcription Elongation and its Control. *Cell*, 120(2):183–193, 2005.

Bell, S. D., Brinkman, A. B., van der Oost, J., and Jackson, S. P. The archaeal TFIIEα homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO reports*, 2(2):133–138, 2001.

Belogurov, G. A., Vassylyeva, M. N., Svetlov, V., Klyuyev, S., Grishin, N. V., Vassylyev, D. G., and Artsimovitch, I. Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. *Molecular Cell*, 26(1):117–129, 2007.

Belogurov, G. A., Mooney, R. A., Svetlov, V., Landick, R., and Artsimovitch, I. Functional specialization of transcription elongation factors. *The EMBO journal*, 28(2):112–12, 2008.

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science*, 334(6062): 1524–1529, 2011.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.

Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R., and Taylor, I. A. Structure of a *Mycobacterium tuberculosis* NusA–RNA complex. *The EMBO journal*, 24(20):3576–3587, 2005.

Beuth, B., García-Mayoral, M. F., Taylor, I. A., and Ramos, A. Scaffold-Independent Analysis of RNA-Protein Interactions: The Nova-1 KH3-RNA Complex. *Journal of the American Chemical Society*, 129(33):10205–10210, 2007.

Bremer, H., Dennis, P. P., et al. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. *Escherichia coli and Salmonella: cellular and molecular biology*, 2:1553–1569, 1996.

Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., and Forterre, P. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biology*, 6(5):R42, 2005.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology*, 6 (3):245–252, 2008.

Browning, D. F. and Busby, S. J. The Regulation of Bacterial Transcription Initiation. *Nature Reviews Microbiology*, 2(1):57–65, 2004.

Bubunenko, M., Court, D. L., Al Refaii, A., Saxena, S., Korepanov, A., Friedman, D. I., Gottesman, M. E., and Alix, J.-H. Nus Transcription Elongation Factors and RNase III Modulate Small Ribosome Subunit Biogenesis in *Escherichia coli*. *Molecular Microbiology*, 87(2):1365–295, 2013.

Buck, M., Gallegos, M.-T., Studholme, D. J., Guo, Y., and Gralla, J. D. The Bacterial Enhancer-dependent ς^{54} (ς^{N}) Transcription Factor. *Journal of bacteriology*, 182(15):4129–4136, 2000.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., et al. Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. *Science*, 273(5258):1058–1073, 1996.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, 56(4):549–561, 1989.

Burmann, B., Scheckenhofer, U., Schweimer, K., and Rösch, P. Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. *Biochem. J*, 435:783–789, 2011.

Burmann, B. M., Luo, X., Rösch, P., Wahl, M. C., and Gottesman, M. E. Fine tuning of the *E. coli* NusB:NusE complex affinity to *BoxA* RNA is required for processive antitermination. *Nucleic acids research*, 38(1):314–326, 2010a.

Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E., and Rösch, P. A NusE: NusG Complex Links Transcription and Translation. *Science*, 328(5977): 501–504, 2010b.

Burmann, B. M., Knauer, S. H., Sevostyanova, A., Schweimer, K., Mooney, R. A., Landick, R., Artsimovitch, I., and Rösch, P. An α Helix to β Barrel Domain Switch Transforms the Transcription Factor RfaH into a Translation Factor. *Cell*, 150(2):291–303, 2012.

Burova, E., Hung, S., Sagitov, V., Stitt, B., and Gottesman, M. *Escherichia coli* NusG Protein Stimulates Transcription Elongation Rates In Vivo and In Vitro. *Journal of bacteriology*, 177 (5):1388–1392, 1995.

Bushnell, D. A., Westover, K. D., Davis, R. E., and Kornberg, R. D. Structural Basis of Transcription: an RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms. *Science*, 303(5660): 983–988, 2004.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell*, 104(6):901–912, 2001.

Campbell, E. A., Westblade, L. F., and Darst, S. A. Regulation of bacterial RNA polymerase σ factor activity: a structural perspective. *Current opinion in microbiology*, 11(2):121–127, 2008.

Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E., and Nudler, E. Termination Factor Rho and its Cofactors NusA and NusG Silence Foreign DNA in *E. coli*. *Science*, 320(5878):935–938, 2008.

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, 407(6802):440–348, 2000.

Castro-Roa, D. and Zenkin, N. *In vitro* experimental system for analysis of transcription–translation coupling. *Nucleic acids research*, 40(6):139–193, 2012.

Cavicchioli, R. Cold-adapted archaea. *Nature Reviews Microbiology*, 4(5):331–343, 2006.

Chaban, B., Ng, S. Y., and Jarrell, K. F. Archaeal habitats – from the extreme to the ordinary. *Canadian journal of microbiology*, 52(2):73–116, 2006.

Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J., Weiss, S., and Ebright, R. H. Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science*, 337 (6094):591–595, 2012.

Chan, C. L. and Landick, R. Dissection of the *his* Leader Pause Site by Base Substitution Reveals a Multipartite Signal that Includes a Pause RNA Hairpin. *Journal of molecular biology*, 233 (1):25–42, 1993.

Chan, C. and Landick, R. The *Salmonella typhimurium his* Operon Leader Region Contains an RNA Hairpin-dependent Transcription Pause Site. Mechanistic implications of the effect on pausing of altered RNA hairpins. *Journal of Biological Chemistry*, 265(34):20796–20804, 1989.

Chatzidaki-Livanis, M., Coyne, M. J., and Comstock, L. E. A Family of Transcriptional Antitermination Factors Necessary for Synthesis of the Capsular Polysaccharides of *Bacteroides fragilis*. *Journal of bacteriology*, 191(23):7288–7295, 2009.

Chen, Y., Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., Hisatake, K., and Handa, H. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes & development*, 23(23):2765–2777, 2009.

Cheng, B. and Price, D. H. Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic acids research*, 36(20): e135–e135, 2008.

Cleveland, D. W., Fischer, S. G., Kirschner, M. W., and Laemmli, U. K. Peptide Mapping by Limited Proteolysis in Sodium Dodecyl Sulfate and Analysis by Gel Electrophoresis. *Journal of Biological Chemistry*, 252(3):1102–1106, 1977.

Condon, C., Squires, C., and Squires, C. L. Control of rRNA Transcription in *Escherichia coli*. *Microbiological reviews*, 59(4):623–645, 1995.

Core, L. J., Waterfall, J. J., and Lis, J. T. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, 322(5909):1845–1848, 2008.

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G., Dengl, S., Geiger, S., Jasiak, A., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J., and Vannini, A. Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.*, 37:337–352, 2008.

Cramer, P., Bushnell, D. A., and Kornberg, R. D. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science*, 292(5523):1863–1876, 2001.

Deighan, P., Diez, C. M., Leibman, M., Hochschild, A., and Nickels, B. E. The bacteriophage λ Q antiterminator protein contacts the β-flap domain of RNA polymerase. *Proceedings of the National Academy of Sciences*, 105(40):15305–15310, 2008.

DeVito, J. and Das, A. Control of transcription processivity in phage lambda: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. *Proceedings of the National Academy of Sciences*, 91(18):8660–8664, 1994.

Dodd, I. B., Shearwin, K. E., and Egan, J. B. Revisited gene regulation in bacteriophage λ. *Current opinion in genetics & development*, 15(2):145–152, 2005.

Drögemüller, J., Stegmann, C. M., Mandal, A., Steiner, T., Burmann, B. M., Gottesman, M. E., Wöhrl, B. M., Rösch, P., Wahl, M. C., and Schweimer, K. An Autoinhibited State in the Structure of *Thermotoga maritima* NusG. *Structure*, 21(3):365–375, 2013.

Duhr, S. and Braun, D. Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences*, 103(52):19678–19682, 2006.

Edwards, T. E. and Sigurdsson, S. T. Site-specific incorporation of nitroxide spin-labels into 2'-positions of nucleic acids. *Nature protocols*, 2(8):1954–1962, 2007.

Edwards, T. E., Okonogi, T. M., Robinson, B. H., and Sigurdsson, S. T. Site-Specific Incorporation of Nitroxide Spin-Labels into Internal Sites of the TAR RNA; Structure-Dependent Dynamics of RNA by EPR Spectroscopy. *Journal of the American Chemical Society*, 123(7): 1527–1528, 2001.

Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I., et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences*, 105(23): 8102–8107, 2008.

Epshtein, V. and Nudler, E. Cooperation Between RNA Polymerase Molecules in Transcription Elongation. *Science*, 300(5620):801–805, 2003.

Epshtein, V., Toulmé, F., Rahmouni, A. R., Borukhov, S., and Nudler, E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J*, 22(18):4719–4727, 2003.

Epshtein, V., Dutta, D., Wade, J., and Nudler, E. An allosteric mechanism of Rho-dependent transcription termination. *Nature*, 463(7278):245–249, 2010.

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., and Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, 357(2):289 – 298, 2006.

Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. Probing protein structure by limited proteolysis. *ACTA BIOCHIMICA POLONICA-ENGLISH EDITION*, 51(299–322), 2004.

Forbes, A. J., Patrie, S. M., Taylor, G. K., Kim, Y.-B., Jiang, L., and Kelleher, N. L. Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2678–2683, 2004.

Forget, D., Langelier, M.-F., Thérien, C., Trinh, V., and Coulombe, B. Photo-Cross-Linking of a Purified Preinitiation Complex Reveals Central Roles for the RNA Polymerase II Mobile Clamp and TFIIE in Initiation Mechanisms. *Molecular and cellular biology*, 24(3):1122–1131, 2004.

Fouqueau, T., Zeller, M. E., Cheung, A. C., Cramer, P., and Thomm, M. The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Research*, page gkt433, 2013.

French, S. L., Santangelo, T. J., Beyer, A. L., and Reeve, J. N. Transcription and Translation are Coupled in Archaea. *Molecular Biology and Evolution*, 24(4):893–895, 2007.

Friedman, D. I., Schauer, A. T., Baumann, M. R., Baron, L. S., and Adhya, S. L. Evidence that ribosomal protein S10 participates in control of transcription termination. *Proceedings of the National Academy of Sciences*, 78(2):1115–1118, 1981.

Friedman, D. I., Baumann, M., and Baron, L. Cooperative Effects of Bacterial Mutations Affecting λ N Gene Expression: I. Isolation and Characterization of a *nusB* Mutant. *Virology*, 73(1):119–127, 1976.

Gambacorta, A., Gliozzi, A., and De Rosa, M. Archaeal lipids and their biotechnological applications. *World Journal of Microbiology and Biotechnology*, 11(1):115–131, 1995.

Ghavi-Helm, Y., Michaut, M., Acker, J., Aude, J.-C., Thuriaux, P., Werner, M., and Soutourina, J. Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. *Genes & development*, 22(14):1934–1947, 2008.

Gilchrist, D. A., Nechaev, S., Lee, C., Ghosh, S. K. B., Collins, J. B., Li, L., Gilmour, D. S., and Adelman, K. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & development*, 22(14):1921–1933, 2008.

Giometti, C. S., Reich, C., Tollaksen, S., Babnigg, G., Lim, H., Zhu, W., Yates III, J., and Olsen, G. Global analysis of a "simple" proteome:*Methanococcus jannaschii*. *Journal of Chromatography B*, 782(1):227–243, 2002.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. *Science*, 292 (5523):1876–1882, 2001.

Gopal, B., Haire, L. F., Gamblin, S. J., Dodson, E. J., Lane, A. N., Papavinasasundaram, K., Colston, M. J., and Dodson, G. Crystal Structure of the Transcription Elongation/Antitermination Factor NusA from *Mycobacterium tuberculosis* at 1.7 Å Resolution. *Journal of molecular biology*, 314(5):1087–1095, 2001.

Greenblatt, J. and Li, J. The *nus*A Gene Protein of *Escherichia coli*: Its Identification and a Demonstration that it Interacts with the Gene N Transcription Anti-termination Protein of Bacteriophage Lambda. *Journal of molecular biology*, 147(1):11–23, 1981.

Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P., and Brochier-Armanet, C. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nature Reviews Microbiology*, 8(101743–752), 2010.

Grohmann, D., Hirtreiter, A., and Werner, F. RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly *in vitro*. *Biochem. J*, 421:339–343, 2009.

Grohmann, D., Klose, D., Klare, J. P., Kay, C. W., Steinhoff, H.-J., and Werner, F. RNA-Binding to Archaeal RNA Polymerase Subunits F/E: A DEER and FRET Study. *Journal of the American Chemical Society*, 132(17):5954–5955, 2010.

Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., Michaelis, J., and Werner, F. The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp During Transcription Initiation and Elongation. *2011*, 43(2):263–274, 2011.

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. The Vienna RNA Websuite. *Nucleic acids research*, 36:W70–W74, 2008.

Grummt, I. Regulation of Mammalian Ribosomal Gene Transcription by RNA Polymerase I. *Progress in nucleic acid research and molecular biology*, 62:109–154, 1998.

Grummt, I. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes & development*, 17(14):1691–1702, 2003.

Grünberg, S., Reich, C., Zeller, M. E., Bartlett, M. S., and Thomm, M. Rearrangement of the RNA polymerase subunit H and the lower jaw in archaeal elongation complexes. *Nucleic acids research*, 38(6):1950–1963, 2010.

Grünberg, S., Warfield, L., and Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature structural & molecular biology*, 19(8):788–796, 2012.

Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A., Teng, M., and Niu, L. Core Structure of the Yeast Spt4-Spt5 Complex: A Conserved Module for Regulation of Transcription Elongation. *Structure*, 16(11):1649–1658, 2008.

Gusarov, I., Nudler, E., et al. Control of Intrinsic Transcription Termination by N and NusA: The Basic Mechanisms. *Cell*, 107(4):437, 2001.

Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R., and Olsen, G. J. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely

thermophilic *Methanococcus* species. *Proceedings of the National Academy of Sciences*, 96(7): 3578–3583, 1999.

Harris, J. K., Kelley, S. T., Spiegelman, G. B., and Pace, N. R. The Genetic Core of the Universal Ancestor. *Genome research*, 3(13):407–412, 2003.

Heckman, K. L. and Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protocols*, 2(4):924–932, 2007.

Heinrich, T., Condon, C., Pfeiffer, T., and Hartmann, R. K. Point Mutations in the Leader boxA of a Plasmid-Encoded *Escherichia coli rrnB* Operon Cause Defective Antitermination In Vivo. *Journal of bacteriology*, 177(13):3793–3800, 1995.

Herr, A., Jensen, M., Dalmay, T., and Baulcombe, D. RNA Polymerase IV Directs Silencing of Endogenous DNA. *Science*, 308(5718):118–120, 2005.

Hirata, A., Klein, B. J., and Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature*, 451(7180):851–854, 2008.

Hirtreiter, A., Damsma, G. E., Cheung, A. C., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C., Cramer, P., and Werner, F. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic acids research*, 38(12):4040–4051, 2010a.

Hirtreiter, A., Grohmann, D., and Werner, F. Molecular mechanisms of RNA polymerase—the F/E (RPB4/7) complex is required for high processivity *in vitro*. *Nucleic acids research*, 38 (2):585–596, 2010b.

Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R., and Ramos, A. KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic acids research*, 40(14):6873–6886, 2012.

Holstege, F., Tantin, D., Carey, M., Van der Vliet, P., and Timmers, H. The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *The EMBO journal*, 14(4):810–819, 1995.

Hubbard, S. J. The structural aspects of limited proteolysis of native proteins. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1382(2):191–206, 1998.

Hubbell, W. L. and Altenbach, C. Investigation of structure and dynamics in membrane proteins using site-directed spin labeling. *Current Opinion in Structural Biology*, 4(4):566–573, 1994.

Hubbell, W. L., Gross, A., Langen, R., and Lietzow, M. A. Recent advances in site-directed spin labeling of proteins. *Current opinion in structural biology*, 8(5):649–656, 1998.

Hubbell, W. L., Cafiso, D. S., and Altenbach, C. Identifying conformational changes with site-directed spin labeling. *Nature Structural & Molecular Biology*, 7(9):735–739, 2000.

Ingham, C. J., Dennis, J., and Furneaux, P. A. Autogenous regulation of transcription termination factor Rho and the requirement for Nus factors in *Bacillus subtilis*. *Molecular microbiology*, 31(2):651–663, 1999.

Ivanov, D., Kwak, Y. T., Guo, J., and Gaynor, R. B. Domains in the SPT5 Protein That Modulate Its Transcriptional Regulatory Properties. *Molecular and cellular biology*, 20(9): 2970–2983, 2000.

Ivanovska, I., Jacques, P.-É., Rando, O. J., Robert, F., and Winston, F. Control of Chromatin Structure by Spt6: Different Consequences in Coding and Regulatory Regions. *Molecular and cellular biology*, 31(3):531–541, 2011.

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., and Duhr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *Assay and drug development technologies*, 9(4):342–353, 2011.

Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Banham, J., Timmel, C., Hilger, D., and Jung, H. DeerAnalysis2006—a Comprehensive Software Package for Analyzing Pulsed ELDOR Data. *Applied Magnetic Resonance*, 30(3):473–498, 2006.

Jeschke, G. and Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Physical Chemistry Chemical Physics*, 9(16):1895–1910, 2007.

Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. Genome-wide Mapping of in Vivo Protein-Dna Interactions. *Science*, 316(5830):1497–1502, 2007.

Kalyani, B. S., Muteeb, G., Qayyum, M. Z., and Sen, R. Interaction with the Nascent RNA is a Prerequisite for the Recruitment of Rho to the Transcription Elongation Complex *In Vitro*. *Journal of molecular biology*, 413(3):548–560, 2011.

Kanno, T., Huettel, B., Mette, M. F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D. P., Matzke, M., and Matzke, A. J. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nature genetics*, 37(7):761–765, 2005.

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., and Ebright, R. H. Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science*, 314(5802):1144–1147, 2006.

Kaplan, C. D., Larsson, K.-M., and Kornberg, R. D. The RNA Polymerase II Trigger Loop Functions in Substrate Selection and Is Directly Targeted by α-Amanitin. *Molecular Cell*, 30 (5):547–556, 2008.

Kashefi, K. and Lovley, D. R. Extending the Upper Temperature Limit for Life. *Science*, 301 (5635):934–934, 2003.

Kettenberger, H., Armache, K.-J., and Cramer, P. Architecture of the RNA polymerase II-TFIIS Complex and Implications for mRNA Cleavage. *Cell*, 114(3):347–357, 2003.

Kim, J. L., Nikolov, D. B., Burley, S. K., et al. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 365(6446):520–527, 1993a.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016):517–522, 2004.

Kim, Y., Geiger, J. H., Hahn, S., Sigler, P. B., et al. Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365(6446):512–520, 1993b.

Klare, J. P. and Steinhoff, H.-J. Spin labeling EPR. *Photosynthesis research*, 102(2):377–390, 2009.

Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X., and Murakami, K. S. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences*, 108(2):546–550, 2011.

Klose, D., Klare, J. P., Grohmann, D., Kay, C. W., Werner, F., and Steinhoff, H.-J. Simulation vs. Reality: A Comparison of In Silico Distance Predictions with DEER and FRET Measurements. *PloS one*, 7(6):e39492, 2012.

Komissarova, N. and Kashlev, M. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proceedings of the National Academy of Sciences*, 94(5):1755–1760, 1997.

Koonin, E. V., Makarova, K. S., and Elkins, J. G. Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biology direct*, 2(1):38, 2007.

Kostrewa, D., Zeller, M. E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature*, 462 (7271):323–330, 2009.

Koyama, H., Ito, T., Nakanishi, T., Kawamura, N., and Sekimizu, K. Transcription elongation factor S-II maintains transcriptional fidelity and confers oxidative stress resistance. *Genes to Cells*, 8(10):779–788, 2003.

Kuznedelov, K., Korzheva, N., Mustaev, A., and Severinov, K. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA–DNA hybrid length. *The EMBO journal*, 21(6):1369–1378, 2002.

Kyrpides, N. C., Woese, C. R., and Ouzounis, C. A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends in biochemical sciences*, 21(11):426–426, 1996.

Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & development*, 12(1):34–44, 1998.

Landick, R. Transcriptional pausing without backtracking. *Proceedings of the National Academy of Sciences*, 106(22):8797–8798, 2009.

Legault, P., Li, J., Mogridge, J., Kay, L. E., and Greenblatt, J. NMR Structure of the bacteriophage lambda N Peptide/*boxB* RNA Complex: Recognition of a GNRA Fold by an Arginine-Rich Motif. *Cell*, 93(2):289–299, 1998.

Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. Sequence-Specific RNA Binding by a Nova KH Domain: Implications for Paraneoplastic Disease and the Fragile X Syndrome. *Cell*, 100(3):323–332, 2000.

Li, J., Mason, S., and Greenblatt, J. Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes & development*, 7(1): 161–172, 1993.

Li, W., Giles, C., and Li, S. Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Research*, 2014.

Lindstrom, D., Squazzo, S., Muster, N., Burckin, T., Wachter, K., Emigh, C., McCleery, J., Yates, J., and Hartzog, G. Dual Roles for Spt5 in Pre-mRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins. *Molecular and cellular biology*, 23(4):1368–1378, 2003.

Lo, M.-C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical Biochemistry*, 332(1):153–159, 2004.

Luo, X., Hsiao, H.-H., Bubunenko, M., Weber, G., Court, D. L., Gottesman, M. E., Urlaub, H., and Wahl, M. C. Structural and Functional Analysis of the *E. coli* NusB-S10 Transcription Antitermination Complex. *Molecular cell*, 32(6):791–802, 2008.

Lüttgen, H., Robelek, R., Mühlberger, R., Diercks, T., Schuster, S. C., Koéhler, P., Kessler, H., Bacher, A., and Richter, G. Transcriptional Regulation by Antitermination. Interaction of RNA with NusB Protein and NusB/NusE Protein Complex of *Escherichia coli*. *Journal of molecular biology*, 316(4):875–885, 2002.

Mah, T.-F., Li, J., Davidson, A. R., and Greenblatt, J. Functional importance of regions in *Escherichia coli* elongation factor NusA that interact with RNA polymerase, the bacteriophage λ n protein and RNA. *Molecular Microbiology*, 34(3):523–527, 1999.

Malone, E. A., Fassler, J. S., and Winston, F. Molecular and genetic characterization of *SPT4*, a gene important for transcription initiation in *Saccharomyces cerevisiae*. *Molecular and General Genetics MGG*, 237(3):449–459, 1993.

Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., and Cramer, P. Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. *The EMBO journal*, 30(7):1302–1310, 2011.

Mason, S. W. and Greenblatt, J. Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. *Genes & development*, 5(8):1504–1512, 1991.

Mason, S. W., Li, J., and Greenblatt, J. Direct Interaction Between Two *Escherichia coli* Transcription Antitermination Factors, NusB and Ribosomal Protein S10. *Journal of molecular biology*, 223(1):55–66, 1992.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology*, 17(10):1272–1278, 2010.

Meinhart, A., Blobel, J., and Cramer, P. An Extended Winged Helix Domain in General Transcription Factor E/IIEα. *Journal of Biological Chemistry*, 278(48):48267–48274, 2003.

Meka, H., Werner, F., Cordell, S. C., Onesti, S., and Brick, P. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Research*, 33(19):6435–6444, 2005.

Miller, T. L. and Wolin, M. J. Enumeration of *Methanobrevibacter smithii* in Human Feces. *Archives of microbiology*, 131(1):14–18, 1982.

Missra, A. and Gilmour, D. S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the *Drosophila* RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences*, 107(25):11301–11306, 2010.

Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z., and Landick, R. Regulator Trafficking on Bacterial Transcription Units In Vivo. *Molecular Cell*, 33(1):97–108, 2009a.

Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M., and Landick, R. Two Structurally Independent Domains of *E. coli* NusG Create Regulatory Plasticity *via* Distinct Interactions with RNA Polymerase and Regulators. *Journal of molecular biology*, 391(2):341–358, 2009b.

Mukherjee, K. and Chatterji, D. Studies on the ω subunit of *Escherichia coli* RNA polymerase. *European Journal of Biochemistry*, 247(3):884–889, 1997.

Murakami, K. S., Masuda, S., and Darst, S. A. Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 å Resolution. *Science*, 296(5571):1280–1284, 2002.

Nakanishi, T., Shimoaraiso, M., Kubo, T., and Natori, S. Structure-Function Relationship of Yeast S-II in Terms of Stimulation of RNA Polymerase II, Arrest Relief, and Suppression of 6-Azauracil Sensitivity. *Journal of Biological Chemistry*, 270(15):8991–8995, 1995.

Nickels, B. E., Roberts, C. W., Sun, H., Roberts, J. W., and Hochschild, A. The σ^{70} Subunit of RNA Polymerase is Contacted by the λQ Antiterminator during Early Elongation. *Molecular cell*, 10(3):611–622, 2002.

Niesen, F. H., Berglund, H., and Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols*, 2(9):2212–2221, 2007.

Nodwell, J. R. and Greenblatt, J. Recognition of *boxA* Antiterminator RNA by the E. coli Antitermination Factors NusB and Ribosomal Protein S10. *Cell*, 72(2):261–268, 1993.

Okuda, M., Tanaka, A., Arai, Y., Satoh, M., Okamura, H., Nagadoi, A., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. A Novel Zinc Finger Structure in the Large Subunit of Human General Transcription Factor TFIIE. *Journal of Biological Chemistry*, 279(49):51395–51403, 2004.

Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., and Pikaard, C. S. Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation. *Cell*, 120(5):613–622, 2005.

Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W., Darst, S. A., et al. Structure and Function of the Transcription Elongation Factor GreB Bound to Bacterial RNA Polymerasere. *Cell*, 114(3):335–345, 2003.

Ouhammouch, M., Werner, F., Weinzierl, R. O., and Geiduschek, E. P. A Fully Recombinant System for Activator-dependent Archaeal Transcription. *Journal of Biological Chemistry*, 279 (50):51719–51721, 2004.

Paget, M., Helmann, J. D., et al. The σ^{70} family of sigma factors. *Genome Biol*, 4(1):203, 2003.

Polyhach, Y., Bordignon, E., and Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. *Physical Chemistry Chemical Physics*, 13(6):2356–2366, 2011.

Proshkin, S., Rahmouni, A. R., Mironov, A., and Nudler, E. Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science*, 328(5977):504–508, 2010.

Quan, S., Zhang, N., French, S., and Squires, C. L. Transcriptional Polarity in rRNA Operons of *Escherichia coli nusA* and *nusB* Mutant Strains. *Journal of bacteriology*, 187(5):1632–1638, 2005.

Reeder, T. C. and Hawley, D. K. Promoter Proximal Sequences Modulate RNA Polymerase II Elongation by a Novel Mechanism. *Cell*, 87(4):767 – 777, 1996.

Revyakin, A., Liu, C., Ebright, R. H., and Strick, T. R. Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science*, 314(5802):1139–1143, 2006.

Richard, P. and Manley, J. L. Transcription termination by nuclear RNA polymerases. *Genes & development*, 23(11):1247–1269, 2009.

Richardson, J. P. Preventing the Synthesis of Unused Transcripts by Rho Factor. *Cell*, 64(6): 1047 – 1049, 1991.

Ring, B. Z. and Roberts, J. W. Function of a Nontranscribed DNA Strand Site in Transcription Elongation. *Cell*, 78(2):317–324, 1994.

Rougvie, A. and Lis, J. T. Postinitiation Transcriptional Control in *Drosophila melanogaster*. *Molecular and cellular biology*, 10(11):6041–6045, 1990.

Rougvie, A. E., Lis, J. T., et al. The RNA Polymerase II Molecule at the 5' End of the Uninduced *hsp70* Gene of *D. melanogaster* is Transcriptionally Engaged. *Cell*, 54(6):795, 1988.

Russell, J. and Zomerdijk, J. C. The RNA polymerase I transcription machinery. In *Biochemical Society symposium*, number 73, page 203. Europe PMC Funders, 2006.

Sainsbury, S., Niesser, J., and Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493(7432):437–440, 2013.

Santangelo, T. J. and Artsimovitch, I. Termination and antitermination: RNA polymerase runs a stop sign. *Nature Reviews Microbiology*, 9(5):319–329, 2011.

Santangelo, T. J., Reeve, J. N., et al. Archaeal RNA Polymerase is Sensitive to Intrinsic Termination Directed by Transcribed and Remote Sequences. *Journal of molecular biology*, 355(2): 196, 2006.

Santangelo, T. J., Matsumi, R., Atomi, H., Imanaka, T., Reeve, J. N., et al. Polarity in Archaeal Operon Transcription in *Thermococcus kodakaraensis*. *Journal of bacteriology*, 190(6):2244–2248, 2008.

Schärpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S., and Rösch, P. Antitermination in bacteriophage λ. The structure of the N36 peptide-*boxB* rna complex. *European Journal of Biochemistry*, 267(8):2397–2408, 2000.

Schiemann, O. and Prisner, T. F. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Quarterly reviews of biophysics*, 40(1):1–53, 2007.

Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H.-P., and Zillig, W. Picrophilus gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. *Journal of bacteriology*, 177(24):7050–7059, 1995.

Schmidt, M. C. and Chamberlin, M. J. *Nus*A Protein of *Escherichia coli* is an Efficient Transcription Termination Factor for Certain Terminator Sites. *Journal of molecular biology*, 195 (4):809–818, 1987.

Schramm, L. and Hernandez, N. Recruitment of RNA polymerase III to its target promoters. *Genes & development*, 16(20):2593–2620, 2002.

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. D. Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science*, 310 (5749):827–834, 2005.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, 31(13):3381–3385, 2003.

Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R., and Artsimovitch, I. The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Molecular cell*, 43(2):253–262, 2011.

Sevostyanova, A., Svetlov, V., Vassylyev, D. G., and Artsimovitch, I. The elongation factor RfaH and the initiation factor σ bind to the same site on the transcription elongation complex. *Proceedings of the National Academy of Sciences*, 105(3):865–870, 2008.

Shibata, R., Bessho, Y., Shinkai, A., Nishimoto, M., Fusatomi, E., Terada, T., Shirouzu, M., and Yokoyama, S. Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochemical and biophysical research communications*, 355(1):122–128, 2007.

Sigurdsson, S., Dirac-Svejstrup, A. B., and Svejstrup, J. Q. Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Molecular Cell*, 38(2): 202–210, 2010.

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *The EMBO journal*, 22(9):2234–2244, 2003.

Steitz, T. A. Structural biology: A mechanism for all polymerases. *Nature*, 391(6664):231–232, 1998.

Stepanova, E., Lee, J., Ozerova, M., Semenova, E., Datsenko, K., Wanner, B. L., Severinov, K., and Borukhov, S. Analysis of Promoter Targets for *Escherichia coli* Transcription Elongation Factor GreA In Vivo and In Vitro. *Journal of bacteriology*, 189(24):8772–8795, 2007.

Stoltenburg, R., Reinemann, C., and Strehlitz, B. SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular engineering*, 24(4):381–403, 2007.

Story, R. M., Li, H., and Abelson, J. N. Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. *Proceedings of the National Academy of Sciences*, 98(4):1465–1470, 2001.

Sullivan, S. L. and Gottesman, M. E. Requirement for E. coli NusG protein in factor-dependent transcription termination. *Cell*, 68(5):989–994, 1992.

Svetlov, V., Belogurov, G. A., Shabrova, E., Vassylyev, D. G., and Artsimovitch, I. Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic acids research*, 35(17): 5694–5705, 2007.

Tan, L., Wiesler, S., Trzaska, D., Carney, H. C., and Weinzierl, R. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J Biol*, 7(10):40, 2008.

Taura, T., Ueguchi, C., Shiba, K., and Ito, K. Insertional disruption of the *nusB* (*ssyB*) gene leads to cold-sensitive growth of *Escherichia coli* and suppression of the *secY24* mutation. *Molecular and General Genetics MGG*, 234(3):429–432, 1992.

Torres, M., Condon, C., Balada, J.-M., Squires, C., and Squires, C. L. Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *The EMBO journal*, 20(14):3811–3820, 2001.

Toulokhonov, I., Artsimovitch, I., and Landick, R. Allosteric Control of RNA Polymerase by a Site That Contacts Nascent RNA Hairpins. *Science*, 292(5517):730–733, 2001.

Újvári, A. and Luse, D. S. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nature structural & molecular biology*, 13(1):49–54, 2005.

Ule, J., Jensen, K., Mele, A., and Darnell, R. B. CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*, 37(4):376–386, 2005.

Uplekar, S., Rougemont, J., Cole, S. T., and Sala, C. High-resolution transcriptome and genome-wide dynamics of RNA polymerase and NusA in *Mycobacterium tuberculosis*. *41*, 41(2):961–977, 2013.

Vaiana, A. C., Neuweiler, H., Schulz, A., Wolfrum, J., Sauer, M., and Smith, J. C. Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of

Experiment and Computer Simulation. *Journal of the American Chemical Society*, 125(47): 14564–14572, 2003.

Vassylyev, D. G., Sekine, S.-i., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*, 417(6890):712–719, 2002.

Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H., and Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, 448(7150): 157–162, 2007a.

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, 448(7150): 163–168, 2007b.

Viktorovskaya, O. V., Appling, F. D., and Schneider, D. A. Yeast Transcription Elongation Factor Spt5 Associates With RNA Polymerase I and RNA Polymerase II Directly. *Journal of Biological Chemistry*, 286(21):18825–18833, 2011.

Vogel, U. and Jensen, K. F. The RNA Chain Elongation Rate in *Escherichia coli* Depends on the Growth Rate. *Journal of bacteriology*, 176(10):2807–2813, 1994.

Vogel, U. and Jensen, K. F. NusA is Required for Ribosomal Antitermination and for Modulation of the Transcription Elongation Rate of both Antiterminated RNA and mRNA. *Journal of Biological Chemistry*, 272(19):12265–12271, 1997.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & development*, 12(3):343–356, 1998.

Wagner, M., Berkner, S., Ajon, M., Driessen, A. M., Lipps, G., and Albers, S.-V. Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. *Biochemical Society Transactions*, 37(1):97, 2009.

Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*, 127 (5):941–954, 2006.

Ward, R., Keeble, D. J., El-Mkami, H., and Norman, D. G. Distance Determination in Heterogeneous DNA Model Systems by Pulsed EPR. *ChemBioChem*, 8(16):1957–1964, 2007.

Ward, R., Bowman, A., Sozudogru, E., El-Mkami, H., Owen-Hughes, T., and Norman, D. G. EPR Distance Measurements in Deuterated Proteins. *Journal of Magnetic Resonance*, 207 (1):164–167, 2010.

Weinzierl, R. O. The nucleotide addition cycle of RNA polymerase is controlled by two molecular hinges in the Bridge Helix domain. *BMC biology*, 8(1):134, 2010.

Wenzel, S., Martins, B., Rosch, P., and Wohrl, B. Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. *Biochem. J*, 425:373–380, 2010.

Werner, F. Structure and function of archaeal RNA polymerases. *Molecular microbiology*, 65 (6):1395–1404, 2007.

Werner, F. Molecular Mechanisms of Transcription Elongation in Archaea. *Chemical reviews*, 113(11):8331–8349, 2013.

Werner, F. and Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology*, 9(2):85–98, 2011.

Werner, F. and Weinzierl, R. O. A Recombinant RNA Polymerase II-like Enzyme Capable of Promoter-Specific Transcription. *Molecular cell*, 10(3):635–646, 2002.

Werner, F. and Weinzierl, R. O. Direct Modulation of RNA Polymerase Core Functions by Basal Transcription Factors. *Molecular and cellular biology*, 25(18):8344–8355, 2005.

Werner, F., Eloranta, J. J., and Weinzierl, R. O. Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic acids research*, 28(21): 4299–4305, 2000.

West, S., Gromak, N., and Proudfoot, N. J. Human 5'–3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016):522–525, 2004.

Westover, K. D., Bushnell, D. A., and Kornberg, R. D. Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center. *Cell*, 119(4): 481–489, 2004.

Williams, T. A., Foster, P. G., Cox, C. J., and Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, 504(7479):231–236, 2013.

Willis, I. M. RNA polymerase III. *European Journal of Biochemistry*, 212(1):1–11, 2005.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. Structure of the 30S ribosomal subunit. *Nature*, 407(6802):327–339, 2000.

Woese, C. R. and Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11):5088–5090, 1977.

Woese, C. R., Kandler, O., and Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, 87(12):4576–4579, 1990.

Wojtas, M. N., Mogni, M., Millet, O., Bell, S. D., and Abrescia, N. G. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic acids research*, 40 (19):9941–9952, 2012.

Woychik, N. A. and Hampsey, M. The RNA Polymerase II Machinery: Structure Illuminates Function. *Cell*, 108(4):453–463, 2002.

Xia, M., Lunsford, R. D., McDevitt, D., and Iordanescu, S. Rapid Method for the Identification of Essential Genes *Staphylococcus aureus*. *Plasmid*, 42(2):144–149, 1999.

Xie, J., Collart, M., Lemaire, M., Stelzer, G., and Meisterernst, M. A single point mutation in TFIIA suppresses NC2 requirement *in vivo*. *The EMBO journal*, 19(4):672–682, 2000.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats is Critical for Processive Transcription Elongation. *Molecular cell*, 21(2):227–237, 2006.

Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature*, 289:751–758, 1981.

Yarnell, W. S. and Roberts, J. W. The Phage λ Gene Q Transcription Antiterminator Binds DNA in the Late Gene Promoter As It Modifies RNA Polymerase. *Cell*, 69(7):1181–1189, 1992.

Young, R. A. and Steitz, J. A. Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. *Proceedings of the National Academy of Sciences*, 75(8):3593–3597, 1978.

Yuzenkova, Y. and Zenkin, N. Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proceedings of the National Academy of Sciences*, 107(24):10878–10883, 2010.

Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A., and Mustaev, A. Mapping of Catalytic Residues in the RNA Polymerase Active Center. *Science*, 107–108, 1996.

Zhou, H., Liu, Q., Gao, Y., Teng, M., and Niu, L. Crystal structure of NusG N-terminal (NGN) domain from *Methanocaldococcus jannaschii* and its interaction with rpoE. *Proteins: Structure, Function, and Bioinformatics*, 76(4):787–793, 2009.

Zhu, W., Reich, C. I., Olsen, G. J., Giometti, C. S., and Yates, J. R. Shotgun Proteomics of *Methanococcus jannaschii* and Insights into Methanogenesis. *Journal of proteome research*, 3 (3):538–548, 2004.